

Evaluation of NEWS on the Salford Dataset: ACP Baseline Model Report



1 INTRODUCTION

A common practice to gauge patients' physiological condition and acute clinical stability is using basic homeostatic measures like vital signs, haematologic findings, and neurological examinations. Various early warning systems have been proposed that use these measurements to assess whether patients face an imminent risk of deterioration or death [6].

The National Early Warning Score (NEWS), developed in conjunction with the Royal College of Physicians, has been widely implemented across the NHS and by healthcare providers outside the UK [5, pp. 13]. It is a points-based early warning system to identify clinical deterioration in acutely ill patients based on routinely recorded vital signs: pulse, breathing rate, blood pressure, blood oxygen level, level of consciousness, and temperature. The paper-based NEWS allocates points in a weighted manner based on these measurements. The sum of these points, which makes up the patient's aggregate NEWS score, is associated with "triggers" - thresholds that correspond to recommended levels of clinical response and monitoring frequency.

While initially designed to monitor for deterioration in secondary-care, the NEWS has been widely tested and validated across healthcare settings. In its latest revision of the standard, the RCP recommended that the NEWS also be used in emergency departments to "aid the initial assessment of patients, ongoing monitoring and patient triage decisions" [5, pp. 18].

Across studies, the NEWS has been found to perform well in discriminating ward patients at risk of death, critical care admission, or cardiac arrest. The present report introduces the novel Salford dataset, which contains records of adult patients with acute or medical presentation at the Salford Royal NHS Foundation Trust from 2014-2022. We then use this dataset to assess the performance of NEWS in predicting adverse patient outcomes, such as those mentioned above, in the dataset population.

Further, we investigate the performance of NEWS on post-discharge mortality, lengthy hospital stay, and readmission within certain timeframes. Anticipating these outcomes is in keeping with the main purpose of an EWS: to effectively recognise patients whose condition is deteriorating but can be improved by timely medical intervention [6, pp. 4].

2 METHODOLOGY

2.1 Data Collection and Contents

The Salford dataset is collated from EPR records collected in real time from admissions to acute beds between 1st April 2014 and 31st March 2022. Further, the dataset includes administrative data produced in retrospect for financial (tariff) purposes.

Our dataset includes records of patients who received ambulatory emergency care (AEC) or same-day emergency care (SDEC) and were discharged before midnight on the day of admission [4]. AEC, alongside the medical admissions unit (EAU) are the two common entry points for adult patients with general medical emergencies. Exceptions present in our records include certain booked admissions and patients directly transferred and admitted to medical or critical care wards.

As a routine part of admission, the responsible staff member (nurse or support worker) records the patient's vital signs within ~ 30 minutes of arrival. The vital signs that make up the NEWS are measured simultaneously, in a standardised manner, using Dinamap monitors. The readings are manually transcribed into EPR and onto the patient's paper chart. Specifically, this data includes:

- Body temperature (°C)
- Pulse (beats/min)
- Diastolic and systolic blood pressure (mmHg)
- Peripheral oxygen saturation (%)

Further, we record the patient's level of consciousness (AVPU), presence of pain, nausea, or vomiting, whether the patient was receiving oxygen at the time of SpO2 measurement and, if applicable, the oxygen flow-rate and mode of delivery. Once these parameters are inputted into EPR, the NEWS score and component sub-scores are automatically computed for the patient. Blood test results (including VBG), where those were performed, are automatically recorded in the laboratory information management system (LIMS) and subsequently imported into EPR.

Other patient information recorded on admission includes: Patient identifier data such as their unique patient number, age, sex, and address. Admission method (e.g., emergency A&E, emergency GP referral, etc), their presenting complaint, arrival time, and the main A&E diagnosis (following initial assessment).

Information gathered in retrospect includes the date and time of discharge, total length of stay (LOS), the wards

the patient was admitted to (in chronological order) and the LOS per ward. Diagnoses coded using the ICD-10-CM standard are complied by a clinical coding team based on the existing codes and clinical notes recorded in EPR. Procedures and services are also recorded in detail in EPR, and are similarly coded in retrospect using the OPCS-4 standard.

2.2 Outcomes

The outcomes we track are death during hospitalisation, death within 30 days after discharge, unanticipated admission to critical care, or unanticipated readmission to the same hospital. To assess the NEWS, we check for the occurrence of death or critical care admission within 24 hours, 48 hours, or at any point after the patient’s arrival. For readmissions, we consider thresholds of 48 hours, 7 days, or 30 days after discharge.

Where it occurs, patient death during their hospitalisation is identified from EPR. Mortality after discharge (within 30 days) is primarily determined from the regional primary care databases (Salford Integrated Records & Greater Manchester Care Record). Critical care admission is identified from patients being admitted into the hospital’s critical care unit (CCU) or the high-dependency medical unit (HH1M).

2.3 Data Preparation

We perform all data manipulation in Python using the Numpy and Pandas libraries. As they are the only parameters that are transcribed manually into EPR, we check the patients’ vital signs for spurious values against hard ranges (e.g., 0 – 100% for SpO2) as well as soft thresholds that we base on the range of physiologically possible values as determined by professional clinical opinion. For parameters that have a recorded NEWS sub-score but their raw value is missing or marked as spurious, we infer the correct value as the midpoint of the relevant NEWS range. Where a parameter has a valid value but the NEWS sub-score is missing, we compute it in accordance with the specification [5, pp. 28-30]. We use the same specification to compute the final NEWS score for any patients where this is not recorded in the original data. Full details of how each parameter is pre-processed are available in appendix A.

2.4 Data Analysis

We assess the ability of NEWS to discriminate between patients who did and did not suffer an adverse outcome using AUROC (area under the ROC curve). We compute this using the scikit-Learn library. The minimum possible AUROC value is 0.5 and corresponds to a completely random relationship between the score in question and the tracked outcomes. Generally, a score of 0.7 – 0.8 indicates reasonable discrimination and values over 0.8 indicate good discrimination.

We investigate the discriminative performance of both the raw NEWS score and the pre-defined “trigger” thresholds for low, medium, and high clinical risk (NEWS of 0 – 4, 5 – 6, and ≥ 7 respectively) [5, pp. 30].

TABLE 1
Summary of Findings

Outcome	Time	NEWS ≥ 7	ROAUC (95% CI)
Critical Event	24h	51.13%	0.914 (0.896 – 0.932)
	48h	41.24%	0.890 (0.877 – 0.903)
	Any	16.12%	0.769 (0.763 – 0.775)
Mortality	24h	62.01%	0.952 (0.936 – 0.967)
	48h	51.15%	0.938 (0.926 – 0.95)
	Any	18.67%	0.770 (0.762 – 0.777)
Critical Care	24h	26.92%	0.831 (0.787 – 0.874)
	48h	24.64%	0.8 (0.772 – 0.829)
	Any	15.1%	0.767 (0.758 – 0.775)
Long LOS	$\geq 24h$	2.50%	0.602 (0.598 – 0.605)
	$\geq 48h$	2.72%	0.607 (0.604 – 0.61)
	≥ 7 days	3.25%	0.598 (0.594 – 0.602)
30-Day Mor/ty	24h	3.68%	0.655 (0.628 – 0.682)
	48h	3.78%	0.655 (0.634 – 0.677)
	Any	4.28%	0.629 (0.62 – 0.638)
Readmission	2 days	1.67%	0.495 (0.478 – 0.511)
	7 days	1.06%	0.503 (0.482 – 0.501)
	30 days	1.31%	0.523 (0.518 – 0.528)

We measure correlation between ordinal variables (NEWS and LOS) using Spearman’s rank correlation. We compare the means of independent samples of normally distributed variables using Welch’s unequal variances t-test, and we use the Mann-Whitney U test for non-normally distributed variables. We present characteristics of variables as *mean \pm std.*

3 RESULTS

Table 1 summarises our findings on the discriminative ability of NEWS for various outcomes.

3.1 Population Characteristics

Records of 170833 admissions were collated, of which 124162 records, corresponding to 63450 unique patients, had a NEWS score recorded on-admission or computed by us during pre-processing. NEWS scores ranges from 0 – 18 with a median score of 1.0 (IQR=2.0). Key characteristics of the dataset population are given in Table 2.

3.2 Critical Event Occurrence

The original study on the NEWS’ discriminative ability considered death within 24 hours of a given set of observations. We define a “Critical Events” as the occurrence of either critical care admission or death during the hospital spell. Since the vital signs and NEWS in our dataset are measured on admission, we investigate the subsets of hospital spells with total LOS within certain thresholds, such as 24 or 48 hours.

A total of 8826 with recorded NEWS experienced a critical event, comprising 4371 critical care admissions and 5308 deaths. The AUROCs (95% CI) of NEWS for critical events in records with 24h LOS, 48h LOS, or any LOS were respectively 0.914 (0.896 – 0.932), 0.890 (0.877 – 0.903), and

TABLE 2
Salford dataset key characteristics

Parameter	Category	<i>n</i> (%)
Sex	Female	65096 (52.43%)
Admission Method	Emergency A&E	113518 (91.43%)
	Emergency GP Referral	5955 (4.8%)
Admission Ward	EAU	82541 (66.48%)
	AEC	36546 (15.27%)
	Other Medical	2980 (2.4%)
	High-Dependency	2095 (1.69%)
Mortality	No	14768 (92.43%)
	Died During Stay	5308 (4.28%)
	Died within 30 Days	4086 (3.29%)
Critical Care	Yes	4371 (3.52%)
Readmitted (30 days)	Yes	15836 (12.75)
NEWS clinical risk	Low (0 – 4)	118223 (95.1%)
	Medium (5 – 6)	3748 (3.02%)
	High (≥ 7)	2333 (1.88%)

TABLE 3
Distribution comparison dep. on critical event occurrence.

Variable	LOS	Mean (Event)	Mean (\neg Event)	Test	<i>p</i>
Age	24h	71.4 \pm 18.92	57.62 \pm 20.92	Welch	0.0
	48h	70.54 \pm 19.56	58.71 \pm 21.09		0.0
	Any	71.72 \pm 16.85	64.5 \pm 20.62		0.0
NEWS	24h	6.82 \pm 4.38	0.75 \pm 1.07	Mann	0.0
	48h	5.91 \pm 4.18	0.78 \pm 1.1		0.0
	Any	3.51 \pm 3.1	1.04 \pm 1.4		0.0

0.769 (0.763 – 0.775). The ROC curves are given in figure 1. The method for computing the AUROC confidence intervals is given in Appendix B.

Critical event occurrence significantly correlated with the NEWS at the same LOS thresholds: chi-square with $\text{NEWS} \geq 7$: $p < 0.0001$ in all cases. In total, 1423/8826 (16%) of critical event occurrences had $\text{NEWS} \geq 7$. Appendix C examines the occurrence of mortality and critical care admission separately.

Patients who experienced critical events were older (71.72 \pm 16.85 years; Welch’s unequal variances t-test: $p < 0.0001$) and had higher NEWS (3.51 \pm 3.1; Mann-Whitney U test: $p < 0.0001$) compared to patients who did not (mean age 64.5 \pm 20.62 years; NEWS 1.04 \pm 1.4). The mean total LOS was 16.32 \pm 24.25 for patients who experienced critical events and 6.07 \pm 13.62 for those who did not (Mann-Whitney U test: $p < 0.0001$). Table 3 gives the same information across the tested LOS thresholds.

3.3 Length of Stay

The total LOS for the patient was significantly correlated with NEWS (Spearman Rank Correlation: $p < 0.0001$). Median length of stay more than doubled for $\text{NEWS} \geq 7$ compared to < 7 (5.90 and 2.21, respectively).

We further investigate whether NEWS can discriminate stays where the LOS will be greater than a certain amount of time; for example, LOS over one day, meaning the patient needs to be monitored overnight. The AUROCs for $\text{LOS} \geq 24\text{h}$, $\geq 48\text{h}$, and ≥ 7 days were 0.602 (0.598 – 0.605), 0.607 (0.604 – 0.61), and 0.598 (0.594 – 0.602) respectively. LOS at the measured thresholds significantly correlated with NEWS: chi-square with $\text{NEWS} \geq 7$: $p < 0.0001$ in all cases.

3.4 Post-Discharge Mortality

The proportion of 30-day mortality in the dataset records was 3.29% ($n = 4086$) and corresponds to 6.44% of patients. The AUROCs (95% CI) of NEWS for 30-day mortality in records with 24h LOS, 48h LOS, or any LOS were respectively 0.655 (0.628 – 0.682), 0.656 (0.634 – 0.677), and 0.629 (0.62 – 0.638). Post-discharge mortality significantly correlated with NEWS at the same LOS thresholds: chi-square with $\text{NEWS} \geq 7$: $p < 0.0001$ in all cases. In total, 175/4090 (4.28%) of occurrences had $\text{NEWS} \geq 7$.

Patients who died within 30 days were older (77.05 \pm 13.10 years; Welch’s unequal variances t-test: $p < 0.0001$) and had higher NEWS (2.0 \pm 2.11; Mann-Whitney U test: $p < 0.0001$) compared to patients who survived (mean age 64.6 \pm 20.54 years; NEWS 1.19 \pm 1.68). The mean total LOS was 13.49 \pm 18.99 for patients who died after discharge and 6.57 \pm 14.65 for survivors (Mann-Whitney U test: $p < 0.0001$). Table 3 gives the same information across the tested LOS thresholds.

3.5 Readmission

The proportion of 30-day readmissions in the dataset was 11.26% ($n=14003$). The AUROCs (95% CI) of NEWS for readmission within 2 days, 7 days, and 30 days were respectively 0.495 (0.478 – 0.511), 0.491 (0.482 – 0.501), and 0.523 (0.518 – 0.528), indicating the NEWS is no better at predicting readmission than random chance. NEWS was not significantly correlated with readmission at any time threshold: chi-square with $\text{NEWS} \geq 7$: $p > 0.23$ in all cases.

4 DISCUSSION

We have investigated applying the NEWS to a large vital signs dataset from a single acute care trust, mainly consisting of emergency admissions and not considering elective procedures or surgical admissions.

The original study by the NEWSDIG to evaluate the then-newly-proposed NEWS tracked death within 24 hours of readings and reported an AUROC score of 0.89. A subsequent study in emergency departments in the Netherlands gave an AUROC score of 0.768 for 30-day mortality based on the admission NEWS.

This report evaluated the NEWS against the combined outcome of unplanned critical care admission and death. We found it appropriate to consider these outcomes as one, as we expect both to be preceded by deranged physiology

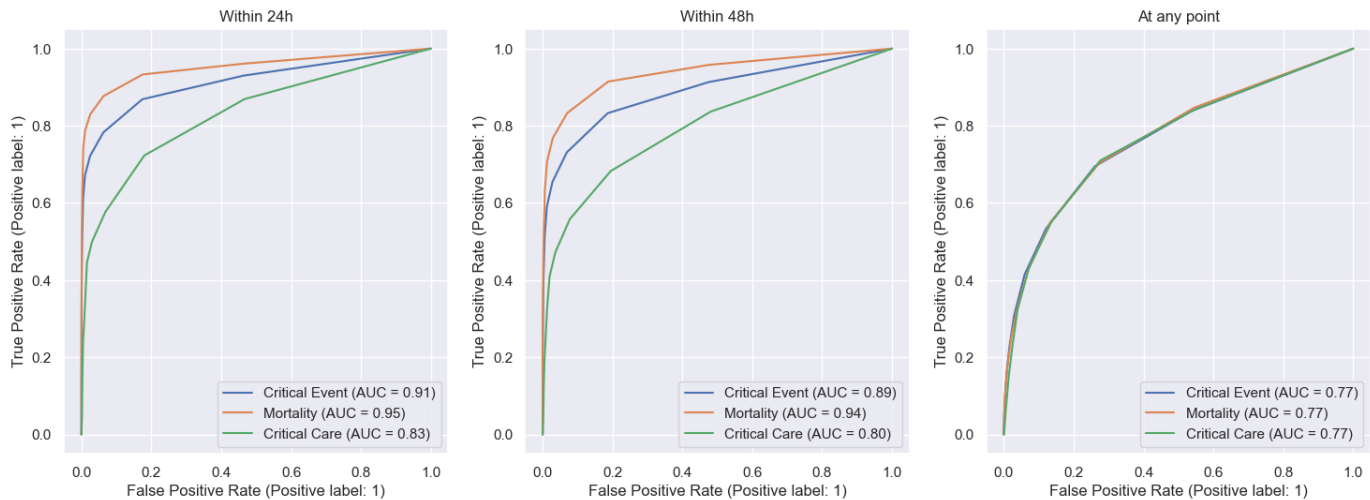


Fig. 1. ROC curves of NEWS for discriminating critical event occurrence in stays with LOS under 24h (left), under 48h (middle), or any LOS (right).

and the clinical response to both events, in terms of urgency and skill, is very similar [6, pp. 4]. We did not track the occurrence of cardiac arrest, as this was not a parameter recorded in our data. However, the foundational NEWS study by [6] concluded this outcome was of limited use as it is largely indistinguishable from mortality except for the fact that a cardiac arrest team has been called in [6, pp. 4].

Our findings regarding the NEWS' discriminative performance are consistent with the previous studies. The NEWS performed best in detecting critical events (in-hospital mortality or admission to critical care). It showed potential in identifying patients at risk of extended hospital stays or post-discharge mortality.

The NEWS was not effective in detecting which patients in our data would be readmitted to the same hospital in the future within any amount of time. Prior work has found that clinical data indicating the severity of the patient's condition, including constructed scores, is useful in predicting 30-day readmission [3, pp. 5]. Dedicated systems for predicting 30-day readmission, such as the HOSPITAL and LACE score, use parameters that the NEWS does not, such as blood test results, procedures the patient underwent, comorbidities and LOS [1].

We showed that parameters present in the data other than vital signs (age) differ in their distribution between patients who experienced adverse outcomes and those who did not. In their comparison of EWS, [6] found that existing EWS systems that considered age showed poor discriminative ability in critical care admission compared to death. The authors noted that the observed performance of an EWS on predicting mortality might not line up with the "true" desired property of an EWS: to effectively recognise patients whose condition is deteriorating but can be improved by timely medical intervention rather than those who will likely not survive [6, pp. 4]. In our view, this makes a strong case for ensuring good discrimination of patients in terms of other adverse outcomes, such as critical care admission, post-discharge mortality, and lengthy hospital stays, to inform and assist clinical decision-making.

In their specification, the NEWSDIG determined it was

unnecessary to apply arbitrary weighting to the score based on the patient's age [5, pp. 19]. Similarly, the NEWS does not consider comorbidities as it is intended to be generic and reflect the physiological perturbations any condition may cause. Given the differing health needs and expected physiological state of patients depending on these excluded factors, we leave the investigation of alternative, stratified early-warning systems as future work.

4.1 Limitations

Our observational dataset is limited to only a single acute care trust. Some measured outcomes and population parameters are inherently variable by hospital. For example, critical care admission may vary between hospitals because each hospital's criteria for admission are different. Making definitive statements about the utility of the NEWS requires further validation in different patient populations and clinical settings.

Patient data were collected primarily by staff under everyday conditions. While this matches the real-life use of the NEWS, operational pressures may hinder the timeliness of data entry and the dataset's reliability. Further, we did not track whether patients had a DNAR order, which can lead to patient death in cases where it otherwise may not have occurred.

A limitation of the dataset was that it only records NEWS measurements taken on admission. To compare with prior studies, we considered adverse event occurrence in hospital stays that lasted under 24 or 48 hours as a proxy for testing NEWS values measured shortly before these events. This practice may induce a bias in our results compared to evaluating the NEWS on data with repeated sets of observations for each patient throughout their hospital stay.

REFERENCES

- [1] Jacques Donzé et al. "Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients: Derivation and Validation of a Prediction Model". In: *JAMA Internal Medicine* 173.8 (Apr. 2013), pp. 632–

638. eprint: https://jamanetwork.com/journals/jamainternalmedicine/articlepdf/1672282/doi120113/_632/_638.pdf.
- [2] J A Hanley and B J McNeil. "The meaning and use of the area under a receiver operating characteristic (ROC) curve." In: *Radiology* 143.1 (1982). PMID: 7063747, pp. 29–36. eprint: <https://doi.org/10.1148/radiology.143.1.7063747>.
 - [3] Elham Mahmoudi et al. "Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review". In: *BMJ* 369 (2020). eprint: <https://www.bmj.com/content/369/bmj.m958.full.pdf>.
 - [4] NHS Improvement and the Ambulatory Emergency Care Network. *Ambulatory emergency care guide: same day emergency care – clinical definition, patient selection and metrics*. England, UK, 2014.
 - [5] Royal College of Physicians. *National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS*. London, 2017.
 - [6] Gary B. Smith et al. "The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death". In: *Resuscitation* 84.4 (2013), pp. 465–470.

APPENDIX A

DATA PRE-PROCESSING

For the components of the NEWS, we set the ranges given in Table 4 and we consider values outside these ranges to be invalid. Where a value is missing or falls outside these ranges, we replace it with a value inferred from the recorded NEWS sub-score. Specifically, we set it to be the midpoint of the relevant NEWS range. Where the sub-score has not been recorded but the raw value is present (and not invalid), we use the value to compute the NEWS sub-score ourselves.

TABLE 4
Valid ranges for NEWS components

Variable	Range	Unit
SpO2	40 – 100	%
Systolic BP	40 – 300	mmHg
Temperature	25 – 45	°C
Pulse	25 – 300	Beats/min
Respiration Rate	5 – 80	Breaths/min

O2 Saturation (SpO2): The NEWS-2 specification gives two scales for this parameter [5, pp. 44]:

- $SpO2_1$: By default.
- $SpO2_2$: For patients with a prescribed oxygen saturation requirement of 88 – 92% (e.g., in patients with hypercapnic respiratory failure).

Since the choice of scale is determined by the responsible clinical staff on a case-by-case basis, we use the following criteria to infer which scale to use when re-computing the NEWS sub-score for SpO2:

- Anyone receiving oxygen using NIV, as recorded directly or in their set of coded procedures ('E85.2').
- Anyone with COPD (as determined by presence of 'J44.*' in their coded diagnoses) AND is receiving oxygen via Venturi 24 or 28.
- Anyone with COPD (as above) and measured SpO2 < 88%.

Finally, if the patient is receiving supplemental oxygen, there is ambiguity as to whether a high O2 NEWS sub-score indicates very high or very low saturation. In that case, we mark the value as missing.

Respiration Rate: In addition to the range given in Table 4, we assume triple-digit values to be erroneous entries of two-digit values (e.g., 250 → 25.0).

Oxygen Flow Rate: This supplemental parameter is entered in mixed units (Litres/min or FiO2). We translate all values to FiO2 where possible:

- Values 1 – 15 are inferred to be in Litres/min.
- Decimal values are inferred to be FiO2, with the exception of 0.5.
- Values of 0.5 and any remaining values are determined based on the device used to deliver the oxygen. Nasal cannula and simple mask correspond to Litres/min, while other devices correspond to FiO2.

We convert Litres/min to FiO2 using the formula $FiO2 = 0.2 + (Litres/min * 4)/100$.

APPENDIX B

AUROC

We assess the discriminative ability of the NEWS using an area under the receiver-operating characteristics (AUROC) curve. This measurement is commonly used to assess the accuracy of a criterion variable whose value will be used to make a binary decision. We construct the ROC curve by plotting the false-positive rate (1 minus the specificity) on the x-axis against the true-positive rate (sensitivity) on the y-axis for each decision threshold for prediction between 0 – 100%. In our methodology, we use scikit-Learn for automatic plotting and AUC computation.

For large samples, the distribution of the AUC is approximately normal. Hence, we use the standard normal distribution to calculate the confidence interval for the AUC as:

$$AUC \pm z_{\alpha/2} SE(AUC) \quad (1)$$

The formula for $SE(AUC)$ is [2]:

$$SE(AUC) = \sqrt{\frac{AUC(1 - AUC) + (N_1 - 1)(Q_1 - AUC^2) + (N_2 - 1)(Q_2 - AUC^2)}{N_1 N_2}} \quad (2)$$

where,

$$Q_1 = \frac{AUC}{2 - AUC}, \quad Q_2 = \frac{2AUC^2}{1 + AUC} \quad (3)$$

APPENDIX C

EXTENDED RESULTS

C.1 Mortality & Critical Care Admission

The proportion of in-hospital mortality in the dataset records was 4.28% ($n = 5328$) and corresponds to 8.386% of patients. The AUROCs (95% CI) of NEWS for mortality in records with 24h LOS, 48h LOS, or any LOS were respectively 0.952 (0.936 – 0.967), 0.938 (0.926 – 0.95), and 0.770 (0.762 – 0.777). Mortality significantly correlated with NEWS at the same LOS thresholds: chi-square with NEWS ≥ 7 : $p < 0.0001$ in all cases. In total, 955/5328 (17.9%) of occurrences had NEWS ≥ 7 .

The proportion of critical care admission in the dataset records was 3.53% ($n = 4384$) and corresponds to 6.90% of patients. The AUROCs (95% CI) of NEWS for critical care admission in records with 24h LOS, 48h LOS, or any LOS were respectively 0.831 (0.787 – 0.874), 0.8 (0.772 – 0.829), and 0.767 (0.758 – 0.775). Admission significantly correlated with NEWS at the same LOS thresholds: chi-square with NEWS ≥ 7 : $p < 0.0001$ in all cases. In total, 662/4384 (15.1%) of occurrences had NEWS ≥ 7 .

C.2 ROC for Length of Stay, Post-Discharge Mortality, and Readmission

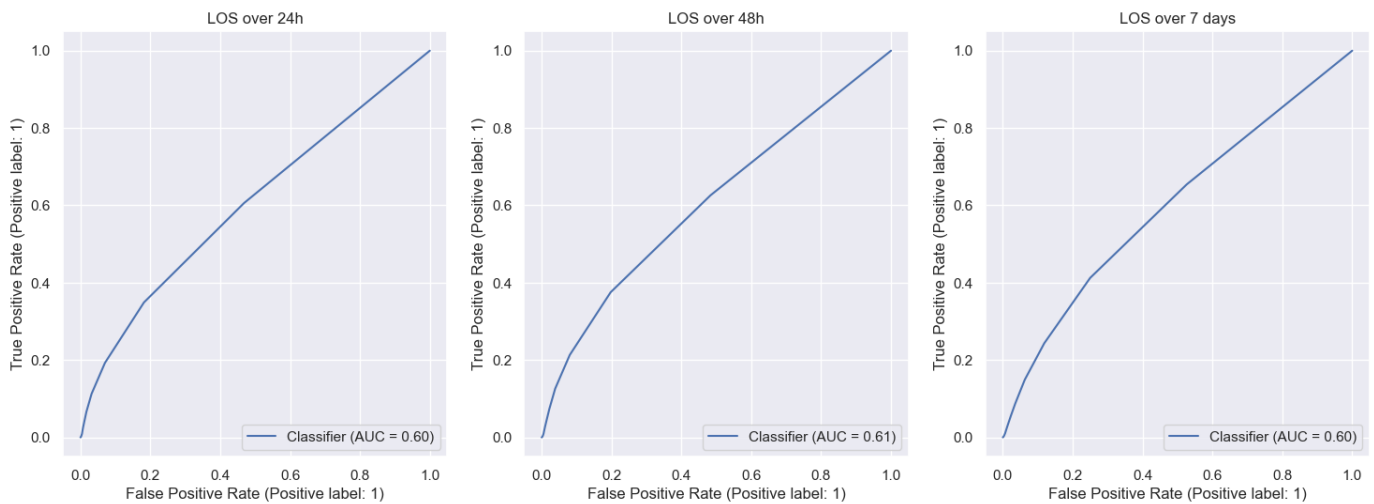


Fig. 2. ROC curves of NEWS for discriminating LOS ≥ 24 h (left), 48h (middle), or 7 days (right).

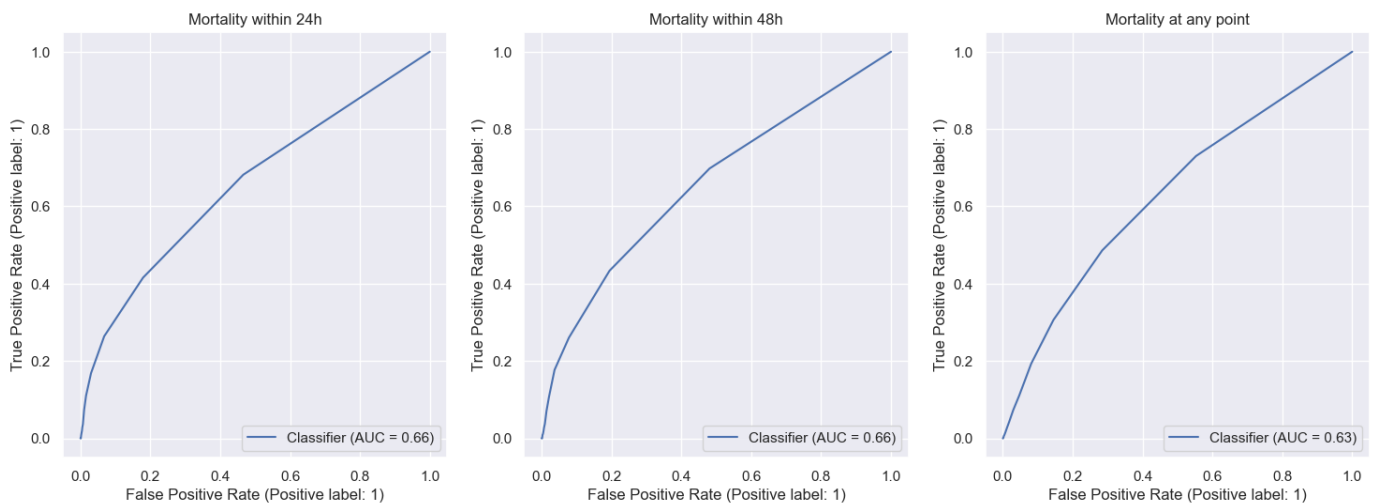


Fig. 3. ROC curves of NEWS for discriminating post-discharge mortality in stays with LOS under 24h (left), under 48h (middle), or any LOS (right).

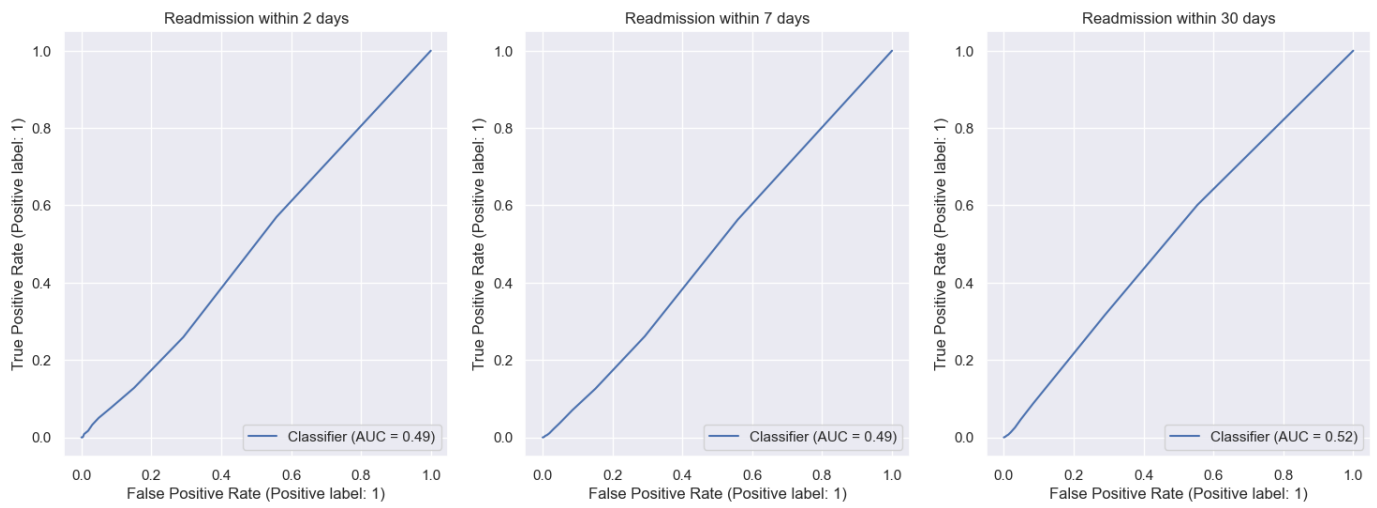


Fig. 4. ROC curves of NEWS for discriminating readmission within 2 days (left), 7 days (middle), or 30 days (right).