

Predicting Acute Clinical Deterioration with Interpretable Machine Learning to support Emergency Care Decision Making

Supplementary Materials

Data Preparation

We filter all manually recorded data features using the ranges in Table 4. Where a value is missing or falls outside these ranges, we replace it with a value inferred from the recorded NEWS2 sub-score. Specifically, we set it to be the midpoint of the relevant NEWS2 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 NEWS2 sub-score ourselves.

Table 4. Valid ranges for manually recorded data features.

Variable	Range	Unit
SpO2	40 – 100	%
Systolic BP	40 – 300	mmHg
Diastolic BP	20 – 200	mmHg
Temperature	25 – 45	°C
Pulse	25 – 300	Beats/min
Respiration Rate	5 – 80	Breaths/min

Furthermore, we apply the following feature-specific filtering:

- **O2 Saturation (SpO2).** The NEWS2 specification gives two scales for this parameter¹:
 - $SpO2_1$: 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 NEWS2 sub-score for SpO2:

- Patients receiving oxygen using NIV, as recorded directly or in their set of coded procedures (OPCS-4 E85.2).
- Patients with COPD (as determined by ICD-10 coded diagnosis J44.*) AND receiving oxygen via Venturi 24 or 28.
- Patients with COPD (as above) and $SpO2 < 88\%$.

Finally, if the patient is receiving supplemental oxygen, there is ambiguity as to whether a high NEWS2 oxygen 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 recorded 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 $FiO_2 = 0.2 + (Litres/min * 4) / 100$.

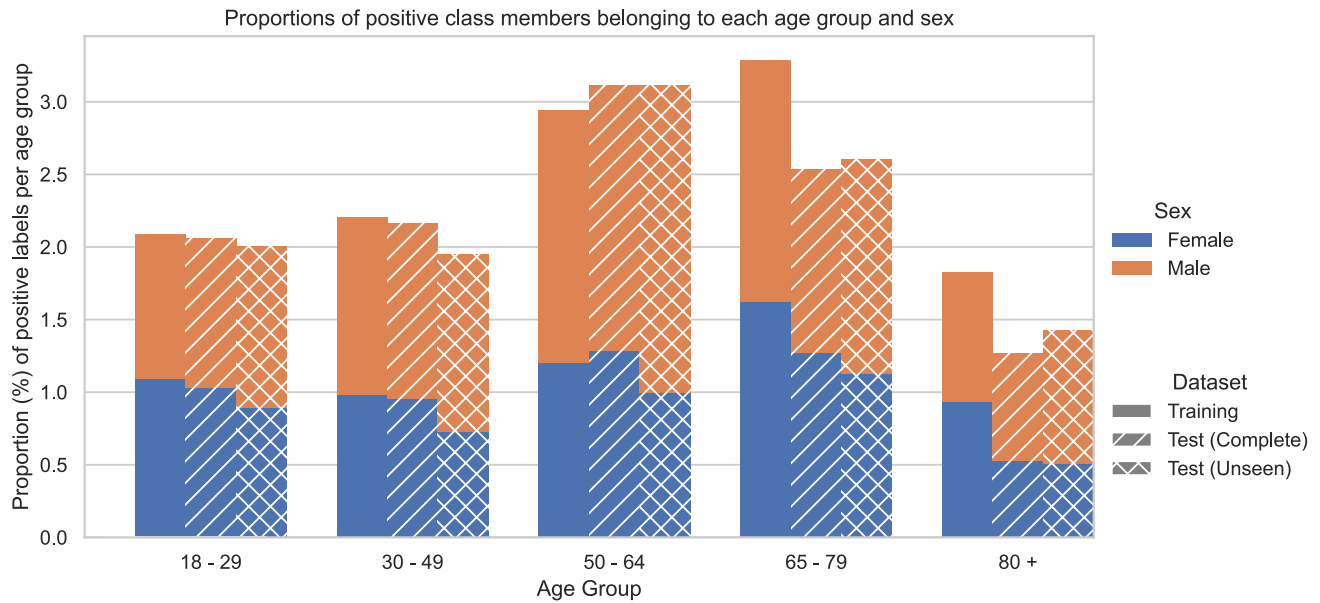


Figure 7. Outcome distribution across age groups and biological sex. Each bar length represents the proportion of positive labels (identified deterioration outcomes) of the corresponding dataset that belonged to the indicated demographic group.

Implementation

Table 5. Hyperparameters for each model across the examined feature sets. These values were chosen using bayesian optimisation after 1000 iterations. Any omitted parameters were assigned their default values.

Model	Parameter	Vitals	& Obs	& Labs	& Notes	& Services
LR	class_weight	None	None	None	None	None
LR-EN	C	0.0592	8.4126	1.1313	8.2228	7.0032
	class_weight	None	None	None	None	None
LR-L1	l1_ratio	0.4133	0.355	0.9356	0.5862	0.8086
	C	4.1169	2.8061	7.6951	4.4909	0.2121
LR-L2	class_weight	None	None	balanced	balanced	None
	C	5.7762	8.6765	9.9825	9.9151	3.5386
LightGBM	class_weight	None	None	None	None	None
	colsample_bytree	0.6075	0.989	0.5526	0.4737	0.4804
	is_unbalance	True	False	False	False	False
	min_child_samples	120	147	73	23	147
	num_leaves	29	63	16	202	13
	reg_alpha	2.6785	0.0402	1.5138	8.8073	0.0001
	reg_lambda	9.9052	0.0009	0.0941	0.0024	9.5277
	scale_pos_weight	None	53	2	4	2
	subsample	0.8714	0.958	0.5702	0.9042	0.4902
LinearSVM	subsample_freq	4	4	1	2	1
	alpha	0.0032	0.0004	0.0002	0.0011	0.0005
XGBoost	class_weight	balanced	None	None	None	None
	alpha	0.0	0.0	0.0	0.0	0.0
	colsample_bytree	0.8389	0.6157	0.6273	0.4414	0.597
	eta	0.0196	0.0454	0.0231	0.0	0.2193
	gamma	0.0	0.0	0.0005	0.0	0.0
	grow_policy	lossguide	lossguide	lossguide	depthwise	lossguide
	lambda	0.0372	0.0019	0.0624	0.157	0.3922
	max_depth	9	7	9	9	3
	min_child_weight	8	9	6	5	4
	scale_pos_weight	3	6	6	15	81
	subsample	0.2505	0.7236	0.4115	0.3563	0.9937
	tree_method	approx	hist	approx	approx	approx

Results

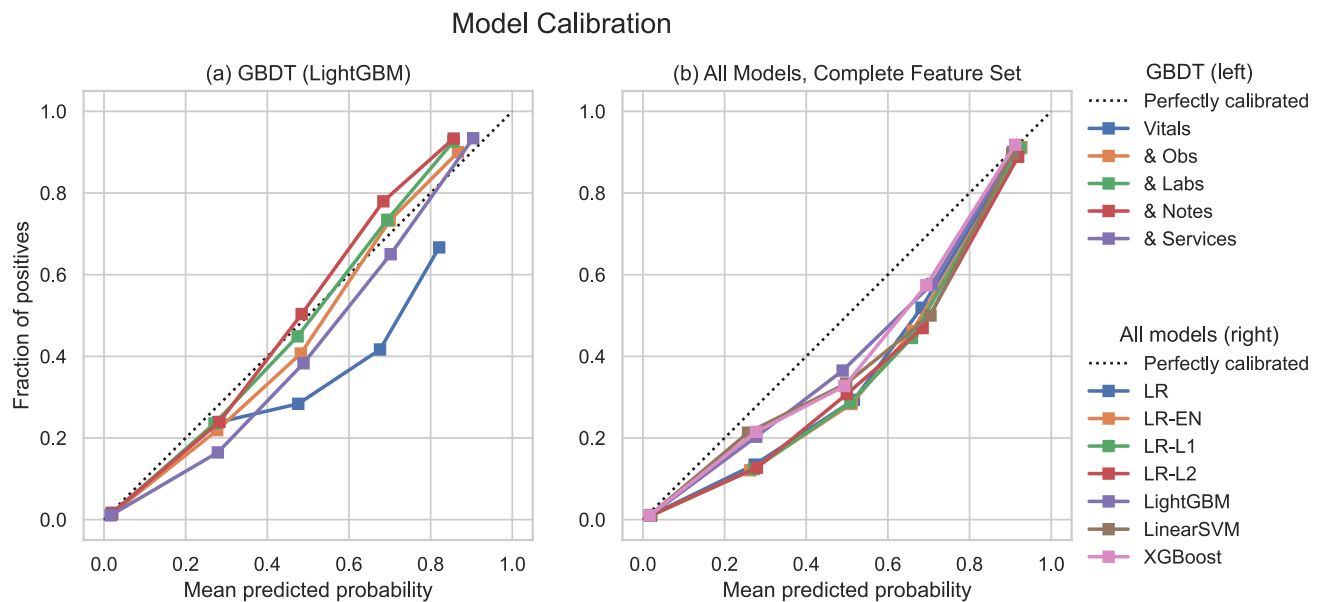


Figure 8. Calibration Curves for **(a)**: GBDT (LightGBM) across feature groups (concatenated incrementally) and **(b)**: All learning algorithms trained the complete feature set (equiv. "& Services"). Curves are plotted from each model's prediction outputs for the complete testing set.

Table 6. Summary of categorical features in the study sample. Each table section (except for ED Diagnosis) presents the 5 most prevalent values for each feature, with the remaining values represented by "Other".

Variable	Value	Total	Training	Test (Complete)	Test (Unseen)
Admission Pathway	Booked Adm.	1807 (1.49%)	899 (1.11%)	908 (2.27%)	582 (2.52%)
	Emg. ED	110627 (91.38%)	74530 (91.89%)	36097 (90.36%)	20461 (88.67%)
	Emg. GP Ref.	5866 (4.85%)	3733 (4.6%)	2133 (5.34%)	1496 (6.48%)
	Emg. OPD	898 (0.74%)	631 (0.78%)	267 (0.67%)	149 (0.65%)
	Non-Emg. Transf.	1180 (0.97%)	792 (0.98%)	388 (0.97%)	279 (1.21%)
	Other	680 (0.56%)	523 (0.64%)	157 (0.39%)	108 (0.47%)
Admission Specialty	ED	18319 (15.13%)	12891 (15.89%)	5428 (13.59%)	3761 (16.3%)
	Acute Int. Med.	91025 (75.19%)	59212 (73.0%)	31813 (79.63%)	17568 (76.13%)
	Gen. Surgery	2612 (2.16%)	2395 (2.95%)	217 (0.54%)	146 (0.63%)
	Nephrology	3454 (2.85%)	2281 (2.81%)	1173 (2.94%)	709 (3.07%)
	Other	4415 (3.65%)	3334 (4.11%)	1081 (2.71%)	745 (3.23%)
	Trauma & Ortho.	1233 (1.02%)	995 (1.23%)	238 (0.6%)	146 (0.63%)
Breathing Device	A - Air	108851 (89.92%)	73233 (90.29%)	35618 (89.16%)	20830 (90.27%)
	N - Nasal cannula	8900 (7.35%)	5547 (6.84%)	3353 (8.39%)	1654 (7.17%)
	Other	1797 (1.48%)	1410 (1.74%)	387 (0.97%)	221 (0.96%)
	RM - Reservoir mask	394 (0.33%)	221 (0.27%)	173 (0.43%)	107 (0.46%)
	SM - Simple mask	685 (0.57%)	308 (0.38%)	377 (0.94%)	249 (1.08%)
	V28 - Venturi 28%	391 (0.32%)	349 (0.43%)	42 (0.11%)	14 (0.06%)
Presenting Complaint	Other	36819 (30.41%)	26077 (32.15%)	10742 (26.89%)	6731 (29.17%)
	abdominal pain	6413 (5.3%)	4514 (5.57%)	1899 (4.75%)	1033 (4.48%)
	chest pain	11585 (9.57%)	7633 (9.41%)	3952 (9.89%)	2371 (10.28%)
	falls	5278 (4.36%)	3421 (4.22%)	1857 (4.65%)	950 (4.12%)
	other	31485 (26.01%)	21565 (26.59%)	9920 (24.83%)	5432 (23.54%)
	shortness of breath	14865 (12.28%)	9337 (11.51%)	5528 (13.84%)	2659 (11.52%)
ED Diagnosis	collaps-	1245 (1.03%)	1245 (1.53%)	0 (0%)	0 (0%)
	confus	356 (0.29%)	356 (0.44%)	0 (0%)	0 (0%)
	cope	592 (0.49%)	348 (0.43%)	244 (0.61%)	150 (0.65%)
	dementia	227 (0.19%)	85 (0.1%)	142 (0.36%)	80 (0.35%)
	diarrh	326 (0.27%)	325 (0.4%)	1 (0.0%)	0 (0%)
	dizz-	89 (0.07%)	89 (0.11%)	0 (0%)	0 (0%)
	fall	1461 (1.21%)	1461 (1.8%)	0 (0%)	0 (0%)
	head	2704 (2.23%)	1992 (2.46%)	712 (1.78%)	434 (1.88%)
	mobility	102 (0.08%)	102 (0.13%)	0 (0%)	0 (0%)
	pain	7162 (5.92%)	7139 (8.8%)	23 (0.06%)	9 (0.04%)
	sudden	1 (0.0%)	1 (0.0%)	0 (0%)	0 (0%)
	tight	5 (0.0%)	5 (0.01%)	0 (0%)	0 (0%)
	vomit	296 (0.24%)	264 (0.33%)	32 (0.08%)	13 (0.06%)

Table 7. AUROC (95% bootstrapped confidence interval) of each classifier type trained on each feature set. 'Test (Complete)' indicates the full testing dataset, 'Test (Unseen)' the testing dataset excluding all patients that had admission records in the training set, and 'Delta' gives the difference (95% bootstrapped confidence interval) in measured performance between the two datasets.

Features	Estimator	Test (Complete)	Test (Unseen)	Delta
Reference	NEWS2	0.79837 (0.78178-0.81774)	0.81211 (0.78853-0.83516)	-0.01374 (-0.02892 - -0.00001)
	LR	0.81202 (0.79589-0.82812)	0.82302 (0.80074-0.84682)	-0.01375 (-0.02613-0.00152)
	LR-EN	0.81084 (0.79487-0.82711)	0.82240 (0.79999-0.84633)	-0.01433 (-0.02630-0.00110)
	LR-L1	0.81202 (0.79587-0.82817)	0.82300 (0.80067-0.84684)	-0.01371 (-0.02613-0.00152)
	LR-L2	0.81204 (0.79593-0.82814)	0.82302 (0.80070-0.84683)	-0.01373 (-0.02611-0.00153)
	LightGBM	0.83451 (0.81795-0.84949)	0.84203 (0.82264-0.86360)	-0.01110 (-0.02066-0.00401)
	LinearSVM	0.80991 (0.79308-0.82680)	0.81609 (0.79223-0.84143)	-0.01150 (-0.02470-0.00489)
	XGBoost	0.83810 (0.82243-0.85295)	0.84792 (0.83019-0.86909)	-0.01432 (-0.02323 - -0.00065)
& Obs	LR	0.83622 (0.82009-0.85170)	0.84324 (0.82577-0.86259)	-0.01140 (-0.02059-0.00670)
	LR-EN	0.83645 (0.82043-0.85198)	0.84346 (0.82612-0.86266)	-0.01140 (-0.02063-0.00673)
	LR-L1	0.83633 (0.82028-0.85191)	0.84337 (0.82579-0.86271)	-0.01144 (-0.02058-0.00676)
	LR-L2	0.83621 (0.82002-0.85168)	0.84324 (0.82580-0.86251)	-0.01139 (-0.02065-0.00666)
	LightGBM	0.86770 (0.85390-0.88193)	0.87881 (0.86276-0.89762)	-0.01402 (-0.02116-0.00076)
	LinearSVM	0.81734 (0.80009-0.83362)	0.80908 (0.78420-0.83397)	-0.01264 (-0.02414-0.00700)
	XGBoost	0.86855 (0.85494-0.88231)	0.87920 (0.86105-0.89896)	-0.01472 (-0.02253 - -0.00123)
& Labs	LR	0.83829 (0.82257-0.85332)	0.84489 (0.82329-0.86387)	-0.01094 (-0.02094-0.00506)
	LR-EN	0.83848 (0.82249-0.85357)	0.84495 (0.82355-0.86405)	-0.01076 (-0.02057-0.00537)
	LR-L1	0.84601 (0.83053-0.86145)	0.85271 (0.83039-0.87151)	-0.01042 (-0.01997-0.00472)
	LR-L2	0.83826 (0.82263-0.85332)	0.84484 (0.82330-0.86376)	-0.01095 (-0.02098-0.00508)
	LightGBM	0.88789 (0.87566-0.90086)	0.89579 (0.87922-0.91310)	-0.01215 (-0.01954-0.00080)
	LinearSVM	0.81036 (0.79234-0.82745)	0.80485 (0.78289-0.82659)	-0.01146 (-0.02057-0.00578)
	XGBoost	0.88099 (0.86808-0.89400)	0.88979 (0.87112-0.90633)	-0.01266 (-0.02039-0.00064)
& Notes	LR	0.84874 (0.83381-0.86303)	0.85299 (0.83578-0.87141)	-0.00879 (-0.01549-0.00595)
	LR-EN	0.85112 (0.83649-0.86548)	0.85363 (0.83651-0.87210)	-0.00705 (-0.01354-0.00785)
	LR-L1	0.85621 (0.84200-0.87038)	0.85986 (0.84446-0.87751)	-0.00734 (-0.01414-0.00693)
	LR-L2	0.85086 (0.83647-0.86516)	0.85360 (0.83637-0.87172)	-0.00726 (-0.01399-0.00744)
	LightGBM	0.89065 (0.87932-0.90209)	0.89731 (0.88408-0.91332)	-0.01079 (-0.01708-0.00224)
	LinearSVM	0.81852 (0.80439-0.83384)	0.83075 (0.81248-0.85447)	-0.01783 (-0.02225 - -0.00194)
	XGBoost	0.88454 (0.87280-0.89649)	0.89101 (0.87656-0.90610)	-0.01075 (-0.01679-0.00086)
& Services	LR	0.89144 (0.87930-0.90370)	0.89280 (0.87805-0.90846)	-0.00693 (-0.01128-0.00637)
	LR-EN	0.89407 (0.88197-0.90623)	0.89555 (0.88032-0.91168)	-0.00696 (-0.01114-0.00591)
	LR-L1	0.89551 (0.88354-0.90677)	0.89778 (0.88336-0.91315)	-0.00729 (-0.01167-0.00505)
	LR-L2	0.89436 (0.88216-0.90637)	0.89591 (0.88078-0.91169)	-0.00701 (-0.01120-0.00562)
	LightGBM	0.92208 (0.91150-0.93140)	0.92295 (0.91214-0.93775)	-0.00648 (-0.01088-0.00560)
	LinearSVM	0.86435 (0.84973-0.87873)	0.85779 (0.84106-0.87776)	-0.00368 (-0.00907-0.01095)
	XGBoost	0.91272 (0.90224-0.92283)	0.91374 (0.90198-0.92932)	-0.00616 (-0.01027-0.00639)

Table 8. Average Precision (95% bootstrapped confidence interval) of each classifier type trained on each feature set. 'Test (Complete)' indicates the full testing dataset, 'Test (Unseen)' the testing dataset excluding all patients that had admission records in the training set, and 'Delta' gives the difference (95% bootstrapped confidence interval) in measured performance between the two datasets.

Features	Estimator	Test (Complete)	Test (Unseen)	Delta
Reference	NEWS2	0.14620 (0.12840-0.16843)	0.16527 (0.13716-0.19425)	-0.01906 (-0.03780-0.00128)
	LR	0.15654 (0.13472-0.18050)	0.17224 (0.14391-0.20818)	-0.01688 (-0.03439-0.00583)
	LR-EN	0.15472 (0.13326-0.17813)	0.16963 (0.14095-0.20585)	-0.01602 (-0.03203-0.00650)
	LR-L1	0.15637 (0.13463-0.18024)	0.17201 (0.14379-0.20773)	-0.01683 (-0.03424-0.00581)
	LR-L2	0.15654 (0.13472-0.18066)	0.17208 (0.14379-0.20789)	-0.01675 (-0.03419-0.00589)
	LightGBM	0.17902 (0.15677-0.20542)	0.20573 (0.17860-0.24097)	-0.02778 (-0.04823 - -0.00317)
	LinearSVM	0.15484 (0.13396-0.17888)	0.17861 (0.14837-0.21558)	-0.01999 (-0.03769-0.00482)
	XGBoost	0.17117 (0.15002-0.19579)	0.19383 (0.16532-0.22807)	-0.02574 (-0.04103 - -0.00242)
Vitals	LR	0.25813 (0.22819-0.28952)	0.30903 (0.27133-0.37279)	-0.05699 (-0.07260 - -0.02352)
	LR-EN	0.25743 (0.22782-0.28908)	0.30870 (0.27093-0.37123)	-0.05701 (-0.07294 - -0.02303)
	LR-L1	0.25673 (0.22718-0.28836)	0.30804 (0.26987-0.37190)	-0.05705 (-0.07247 - -0.02289)
	LR-L2	0.25767 (0.22815-0.28926)	0.30914 (0.27135-0.37176)	-0.05721 (-0.07297 - -0.02316)
	LightGBM	0.31362 (0.28485-0.34723)	0.37001 (0.32539-0.42758)	-0.06763 (-0.08137 - -0.02843)
	LinearSVM	0.25792 (0.22883-0.28772)	0.28442 (0.24230-0.35246)	-0.06619 (-0.08441 - -0.03117)
	XGBoost	0.30981 (0.28000-0.34358)	0.36394 (0.32159-0.41987)	-0.06574 (-0.07730 - -0.02969)
& Obs	LR	0.25956 (0.22983-0.29112)	0.30524 (0.26482-0.37422)	-0.05144 (-0.06715 - -0.02000)
	LR-EN	0.25752 (0.22833-0.28941)	0.30210 (0.26191-0.37002)	-0.05019 (-0.06629 - -0.01838)
	LR-L1	0.24392 (0.21573-0.27720)	0.28401 (0.24618-0.35152)	-0.04440 (-0.06184 - -0.00980)
	LR-L2	0.25893 (0.22947-0.29100)	0.30484 (0.26505-0.37455)	-0.05161 (-0.06791 - -0.01961)
	LightGBM	0.31363 (0.28367-0.34813)	0.36479 (0.31386-0.42422)	-0.06396 (-0.06988 - -0.01894)
	LinearSVM	0.25743 (0.22934-0.28827)	0.25249 (0.20524-0.32339)	-0.05987 (-0.08037 - -0.02942)
	XGBoost	0.31982 (0.28903-0.35317)	0.37818 (0.33927-0.43446)	-0.06876 (-0.07788 - -0.03054)
& Labs	LR	0.26304 (0.23415-0.29418)	0.31080 (0.26920-0.37828)	-0.05432 (-0.06790 - -0.02268)
	LR-EN	0.26395 (0.23514-0.29546)	0.31155 (0.27015-0.37778)	-0.05416 (-0.06802 - -0.02305)
	LR-L1	0.24261 (0.21452-0.27399)	0.28588 (0.24262-0.35254)	-0.04774 (-0.06149 - -0.01234)
	LR-L2	0.26380 (0.23494-0.29479)	0.31169 (0.27030-0.37863)	-0.05444 (-0.06805 - -0.02321)
	LightGBM	0.32338 (0.29322-0.35592)	0.37353 (0.32555-0.42557)	-0.06296 (-0.07043 - -0.02216)
	LinearSVM	0.25459 (0.22619-0.28545)	0.25961 (0.21181-0.32756)	-0.06210 (-0.07710 - -0.02972)
	XGBoost	0.29689 (0.26741-0.32912)	0.34996 (0.31093-0.41418)	-0.06141 (-0.07232 - -0.02160)
& Notes	LR	0.46334 (0.42731-0.49760)	0.47583 (0.41855-0.52883)	-0.03587 (-0.04615-0.00279)
	LR-EN	0.46615 (0.43139-0.50070)	0.47778 (0.41715-0.52816)	-0.03458 (-0.04480-0.00443)
	LR-L1	0.45835 (0.42212-0.49458)	0.46817 (0.40793-0.51891)	-0.03170 (-0.04219-0.00539)
	LR-L2	0.46588 (0.43123-0.50015)	0.47774 (0.41752-0.52838)	-0.03447 (-0.04454-0.00380)
	LightGBM	0.51268 (0.47813-0.54589)	0.52347 (0.46740-0.57071)	-0.03662 (-0.05118-0.00287)
	LinearSVM	0.44708 (0.41128-0.48127)	0.43666 (0.39277-0.48465)	-0.03908 (-0.04888 - -0.00210)
	XGBoost	0.49718 (0.46321-0.53057)	0.50369 (0.45329-0.55213)	-0.03409 (-0.04570-0.00897)
& Services	LR	0.46334 (0.42731-0.49760)	0.47583 (0.41855-0.52883)	-0.03587 (-0.04615-0.00279)
	LR-EN	0.46615 (0.43139-0.50070)	0.47778 (0.41715-0.52816)	-0.03458 (-0.04480-0.00443)
	LR-L1	0.45835 (0.42212-0.49458)	0.46817 (0.40793-0.51891)	-0.03170 (-0.04219-0.00539)
	LR-L2	0.46588 (0.43123-0.50015)	0.47774 (0.41752-0.52838)	-0.03447 (-0.04454-0.00380)
	LightGBM	0.51268 (0.47813-0.54589)	0.52347 (0.46740-0.57071)	-0.03662 (-0.05118-0.00287)
	LinearSVM	0.44708 (0.41128-0.48127)	0.43666 (0.39277-0.48465)	-0.03908 (-0.04888 - -0.00210)
	XGBoost	0.49718 (0.46321-0.53057)	0.50369 (0.45329-0.55213)	-0.03409 (-0.04570-0.00897)

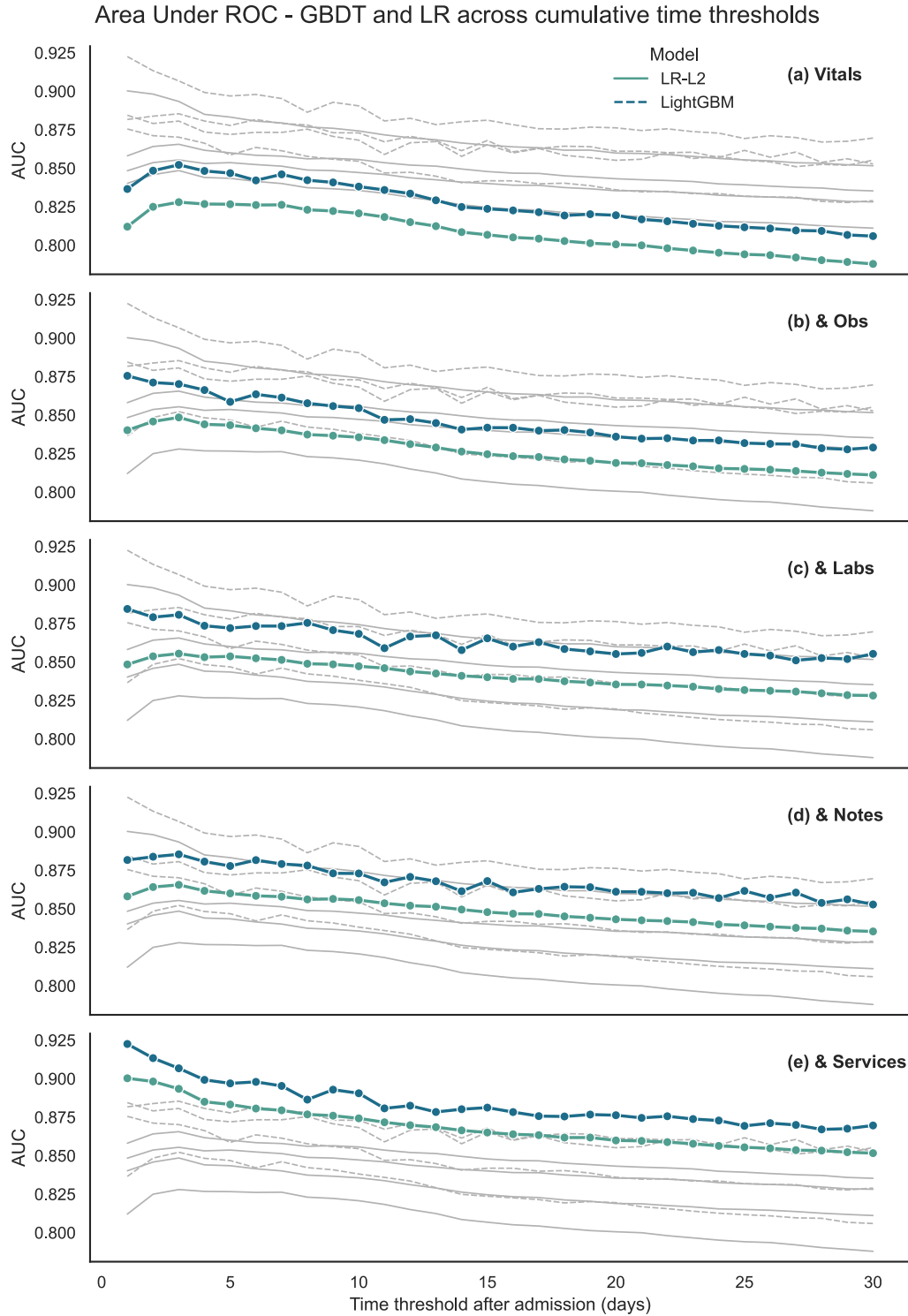


Figure 9. AUROC of logistic regression (L2 penalty) and GBDT (LightGBM) across thresholds for labelling critical deterioration events. Each section (a-e) corresponds to incrementally including the indicated feature sets (from Methods Table 3) as training data. For each one, and for each x-axis value, we train independent models to identify critical deterioration up to the corresponding number of days after admission and measure their AUROC on the validation set (y-axis). The lines in colour represent the performance for the indicated feature set, and the lines in gray represent the lines from the other sections for easier visual comparison.

Table 9. Coefficients of numerical features in logistic regression models. LR: Logistic regression; LR-L1: LR with L1 penalty regularisation; LR-L2: LR with L2 penalty regularisation; LR-EN: LR with Elastic Net regularisation.

	LR	LR-L1	LR-L2	LR-EN
Sent To SDEC	−1.0367	−1.0309	−1.0391	−1.0384
Sodium	−0.8081	−0.7383	−0.8016	−0.8073
Age	−0.546	−0.4507	−0.5434	−0.5433
FiO2	−0.4453	−0.4357	−0.4423	−0.4440
SpO2	−0.1996	−0.1972	−0.1994	−0.1993
Creatinine	−0.1968	−0.1889	−0.1953	−0.1950
Assisted Breathing	−0.1301	−0.1242	−0.1292	−0.1288
30-day Readmission	−0.1131	−0.0779	−0.1144	−0.1143
Temperature	−0.0525	−0.0449	−0.0517	−0.0520
Alert (AVCPU)	−0.0456	−0.0183	−0.0408	−0.0413
Systolic BP	−0.0205	−0.0216	−0.0212	−0.0215
Urea	0.0075	−0.0013	0.0059	0.0060
Pain	0.0766	0.0748	0.0790	0.0785
Nausea	0.0796	0.0407	0.0798	0.0791
Diastolic BP	0.1157	0.0803	0.1114	0.1130
Pulse	0.1193	0.1191	0.1192	0.1191
Haemoglobin	0.1403	0.0989	0.1384	0.1384
Vomiting	0.1891	0.1940	0.1902	0.1899
Respiration Rate	0.1976	0.1984	0.1971	0.1975
Female	0.2662	0.2556	0.2664	0.2670
Lying Down	0.3488	0.3320	0.3484	0.3488
Potassium	0.9721	0.4844	1.4895	0.6155

Categorical Feature Interactions - GBDT (LightGBM)

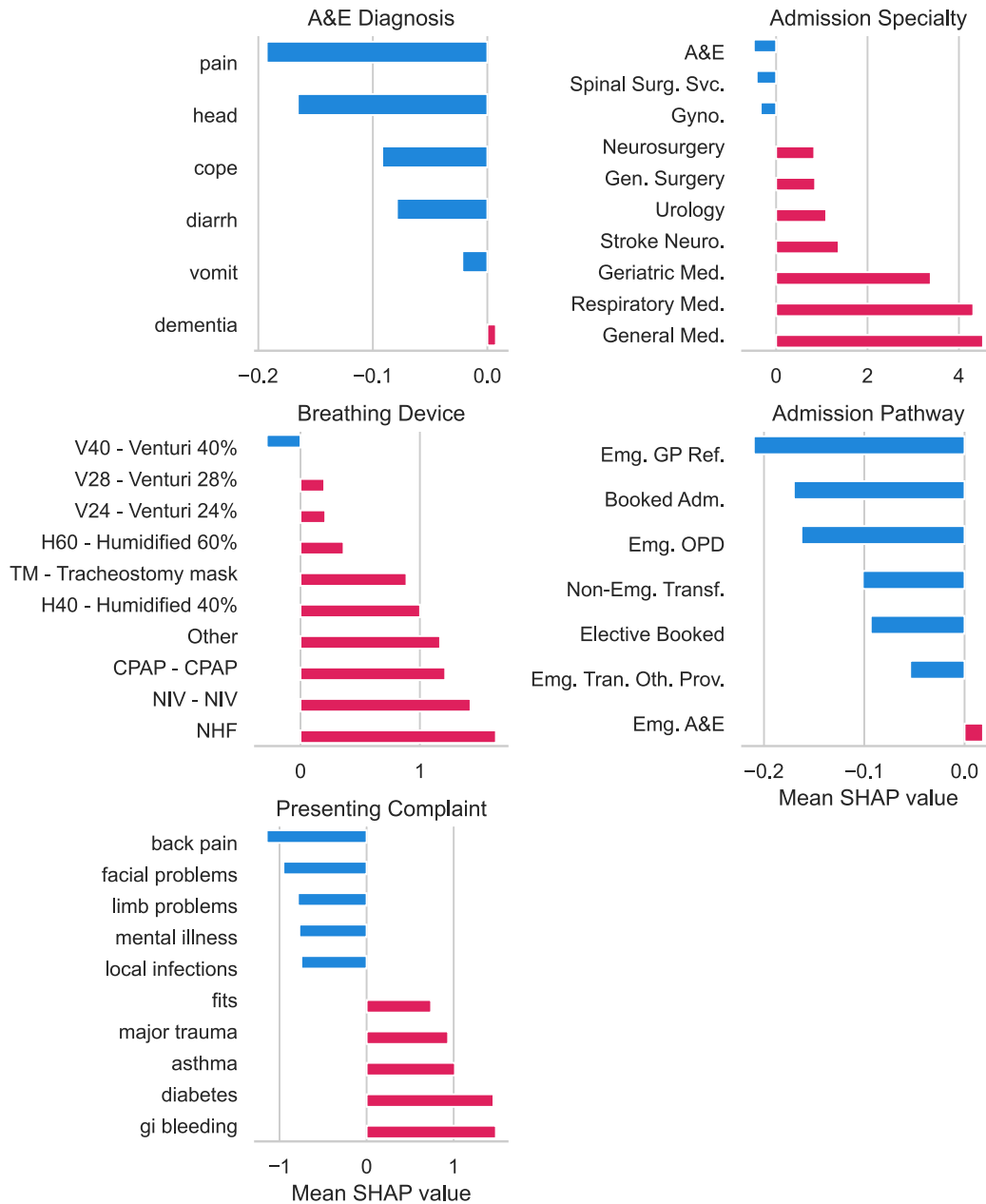


Figure 10. SHAP feature importances for categorical feature values under the GBDT (LightGBM) model. Each bar shows the mean SHAP value for the indicated value of the corresponding categorical feature. Values to the right of $x = 0$ (red) contribute, on average, towards predictions of deterioration, while values to the left (blue) contribute towards negative predictions (no deterioration). We provide the highest-ranked values per feature by mean absolute SHAP value.

Table 10. Coefficients of categorical feature values in logistic regression models. We provide the 5 highest ranked values per feature by mean absolute coefficient over the presented models. LR: Logistic regression; LR-L1: LR with L1 penalty regularisation; LR-L2: LR with L2 penalty regularisation; LR-EN: LR with Elastic Net regularisation.

Feature	Value	LR	LR-EN	LR-L1	LR-L2
ED Diagnosis	dementia	-5.7765	-1.1784	0.0000	-0.9183
	dizz-	-4.5919	-0.5101	0.0000	-0.5284
	Unknown	1.1955	0.5185	0.4662	0.6791
	diarrh	0.9823	0.2567	0.0000	0.4210
	collaps-	0.8968	0.2167	0.0000	0.3797
Admission Pathway	Emg. Tran. Oth. Prov.	1.1180	1.5083	0.8563	1.6927
	Elective Booked	-2.8855	-1.2208	0.0000	-0.8683
	Emg. OPD	-1.7830	-1.2437	0.0000	-0.9072
	Booked Adm.	-1.5348	-1.0707	0.0000	-0.7858
	Emg. ED	0.4351	0.8734	0.9080	1.1078
Admission Specialty	General Med.	6.0295	4.4717	4.5527	4.2806
	Respiratory Med.	5.7150	4.1316	3.7710	3.8547
	Geriatric Med.	5.1367	3.5265	2.4572	3.1878
	Bariatric Surg.	8.3252	2.2719	0.0000	1.7883
	Spinal Surg. Svc.	-6.1272	-2.0428	0.0000	-1.5927
Breathing Device	NHF	3.1081	3.1803	3.0982	3.0330
	Unknown	-5.8139	-1.2859	0.0000	-0.6555
	A - Air	-1.4371	-1.6811	-1.3501	-0.8618
	NIV - NIV	1.1222	1.2162	1.2235	1.1433
	Other	0.8460	0.9446	1.1644	0.8756
Presenting Complaint	ear problems	-6.0519	-1.5286	0.0000	-1.2409
	facial problems	-5.5940	-1.4631	0.0000	-1.1675
	gi bleeding	2.8015	1.7446	1.3005	1.7253
	diabetes	2.7086	1.6517	1.1506	1.6313
	testicular pain	-4.8856	-0.9064	0.0000	-0.8156

Table 11. Differential Fairness Bias Amplification (95% bootstrapped confidence interval) of each classifier type trained on each feature set. The columns "Sex", "Age", and "Sex & Age" indicate the protected characteristic for each measurement - biological sex, age group (per Supplementary Figure 7), or both.

Features	Estimator	Sex	Age	Sex & Age
Vitals	LR	-0.240 (-0.374 - -0.115)	0.107 (0.060 - 0.267)	-0.213 (-0.343 - -0.010)
	LR-EN	-0.240 (-0.373 - -0.114)	0.110 (0.065 - 0.269)	-0.211 (-0.337 - -0.007)
	LR-L1	-0.240 (-0.374 - -0.115)	0.106 (0.060 - 0.267)	-0.214 (-0.344 - -0.011)
	LR-L2	-0.240 (-0.374 - -0.115)	0.107 (0.060 - 0.267)	-0.213 (-0.343 - -0.011)
	LightGBM	-0.245 (-0.396 - -0.123)	0.030 (-0.019 - 0.194)	-0.312 (-0.428 - -0.089)
	LinearSVM	-0.253 (-0.385 - -0.125)	0.200 (0.142 - 0.359)	-0.106 (-0.229 - 0.077)
	XGBoost	-0.242 (-0.392 - -0.125)	0.089 (0.045 - 0.252)	-0.250 (-0.348 - -0.036)
& Obs	LR	0.050 (-0.114 - 0.114)	-0.483 (-0.554 - -0.287)	-0.617 (-0.773 - -0.441)
	LR-EN	0.050 (-0.115 - 0.113)	-0.484 (-0.555 - -0.288)	-0.619 (-0.775 - -0.442)
	LR-L1	0.049 (-0.116 - 0.112)	-0.484 (-0.555 - -0.288)	-0.619 (-0.774 - -0.442)
	LR-L2	0.051 (-0.115 - 0.114)	-0.484 (-0.555 - -0.288)	-0.618 (-0.776 - -0.442)
	LightGBM	0.006 (-0.140 - 0.081)	-0.320 (-0.407 - -0.138)	-0.465 (-0.622 - -0.302)
	LinearSVM	-0.147 (-0.271 - -0.055)	-0.687 (-0.750 - -0.554)	-1.016 (-1.111 - -0.826)
	XGBoost	-0.022 (-0.174 - 0.058)	-0.298 (-0.349 - -0.101)	-0.494 (-0.636 - -0.298)
& Labs	LR	0.074 (-0.098 - 0.143)	-0.500 (-0.564 - -0.312)	-0.604 (-0.769 - -0.431)
	LR-EN	0.074 (-0.099 - 0.141)	-0.499 (-0.562 - -0.311)	-0.602 (-0.771 - -0.431)
	LR-L1	0.095 (-0.079 - 0.168)	-0.504 (-0.552 - -0.314)	-0.589 (-0.685 - -0.351)
	LR-L2	0.074 (-0.098 - 0.143)	-0.499 (-0.564 - -0.312)	-0.602 (-0.768 - -0.430)
	LightGBM	0.034 (-0.123 - 0.111)	-0.317 (-0.388 - -0.144)	-0.416 (-0.583 - -0.265)
	LinearSVM	-0.047 (-0.172 - 0.040)	-0.592 (-0.671 - -0.406)	-0.660 (-0.805 - -0.464)
	XGBoost	-0.015 (-0.165 - 0.067)	-0.329 (-0.404 - -0.140)	-0.514 (-0.648 - -0.324)
& Notes	LR	0.065 (-0.110 - 0.139)	-0.475 (-0.539 - -0.285)	-0.577 (-0.760 - -0.404)
	LR-EN	0.069 (-0.109 - 0.143)	-0.476 (-0.543 - -0.286)	-0.573 (-0.761 - -0.402)
	LR-L1	0.076 (-0.104 - 0.155)	-0.499 (-0.556 - -0.310)	-0.596 (-0.766 - -0.420)
	LR-L2	0.069 (-0.109 - 0.143)	-0.476 (-0.542 - -0.286)	-0.573 (-0.761 - -0.401)
	LightGBM	0.013 (-0.132 - 0.084)	-0.313 (-0.371 - -0.131)	-0.453 (-0.606 - -0.292)
	LinearSVM	-0.047 (-0.187 - 0.055)	-0.657 (-0.733 - -0.514)	-0.885 (-0.996 - -0.694)
	XGBoost	-0.061 (-0.215 - 0.017)	-0.187 (-0.242 - -0.032)	-0.472 (-0.568 - -0.245)
& Services	LR	0.041 (-0.088 - 0.114)	-0.454 (-0.532 - -0.241)	-0.588 (-0.734 - -0.372)
	LR-EN	0.047 (-0.085 - 0.120)	-0.459 (-0.538 - -0.244)	-0.587 (-0.728 - -0.372)
	LR-L1	0.047 (-0.086 - 0.117)	-0.456 (-0.543 - -0.239)	-0.584 (-0.721 - -0.374)
	LR-L2	0.050 (-0.083 - 0.123)	-0.459 (-0.539 - -0.244)	-0.584 (-0.727 - -0.370)
	LightGBM	0.004 (-0.118 - 0.086)	-0.263 (-0.312 - -0.048)	-0.396 (-0.501 - -0.212)
	LinearSVM	-0.137 (-0.229 - -0.043)	-0.579 (-0.647 - -0.399)	-0.878 (-0.959 - -0.627)
	XGBoost	-0.005 (-0.122 - 0.078)	-0.260 (-0.322 - -0.049)	-0.420 (-0.527 - -0.219)

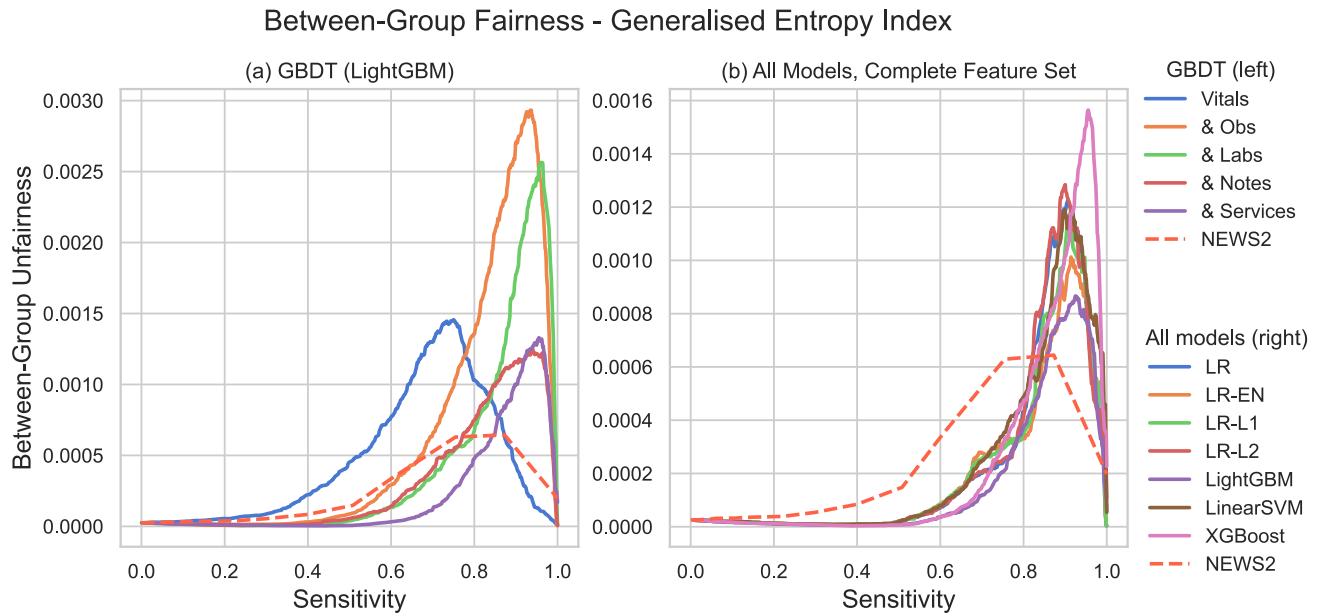


Figure 11. Between-Group Generalised Entropy vs Sensitivity curves of (a): GBDT across the tested feature sets, and (b): All classifier types trained on the complete feature set. We plot the between-group component of the generalised entropy index, representing unfairness between demographic groups defined by the protected characteristics of age group and sex, per Supplementary Figure 7. The remainder of the generalised entropy, as presented in Figure 6 is the *within-group* component, representing all other potential biases. A lower value on the y-axis indicates a more fair distribution of 'benefit', i.e. of receiving a positive prediction, between the demographic groups we consider. A theoretical 'perfect' model would yield a single point (0, 1) in the lower-right corner of the plot.

References

1. Royal College of Physicians. *National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS* (RCP, 2017).