

Table S1– Predictor variables included in logistic stepwise regression analysis used to develop algorithm

Gender	Male / female
Age group (years) (dataset excluded those < 12 and over 80)	12 – 60 / 61+
BMI category	Underweight (0 - < 18.5 kg/m ²) Normal (18.5 - < 25) Overweight (25 - < 30) Obese (>= 30)
Smoking status	Non-smoker / Current smoker / Ex-smoker
Eosinophil count	<= 400 µL ⁻¹ / > 400 µL ⁻¹
Charlson comorbidity Index (CCI) score	0 / 1-4 / 5+
Diagnosis of rhinitis and/or use of drugs	No / yes
Diabetes and /or therapy	No / yes
Heart failure	No / yes
Ischaemic heart disease	No / yes
Anxiety and/or depression	No / yes
History of anaphylaxis	No / yes
Paracetamol prescriptions	No / yes
NSAIDS prescriptions	No / yes
GINA management step	none 1 2 3 4 5
GINA management step (regrouped)	No therapy / Step 1 -2 (baseline) / Step 3 / Step 4 -5
Baseline average SABA daily dosage (mcg)	0 / 1 – 200 / 201 – 400 / 401+
Baseline average ICS daily dosage (mcg)	0 / 1 – 200 / 201 – 400 / 401 - 800 / 801+
GP consultations for LRTI	0 / 1 / 2+
Courses of acute oral steroids	0 / 1 / 2 / 3+
Treatment therapy	None / SABA+-SAMA LABA+-LAMA ICS – LOW DOSE ICS – HIGH DOSE ICS +- LABA +- LAMA – LOW DOSE ICS +- LABA +- LAMA – HIGH DOSE LTRA LTRA +- LABA +- LAMA ICS + LTRA – LOW DOSE ICS + LTRA – HIGH DOSE ICS + LTRA +- LABA +- LAMA – HIGH DOSE OTHER
A & E attendance	0 / 1+
Outpatient department attendance	0 / 1+
Hospitalisation (HES data)	0 / 1+
Outcome: Hospitalisation	0 / 1+

¹ All variables coded no / yes unless otherwise specified. SABA: short acting bronchodilator, ICS: inhaled corticosteroids, LRTI: lower respiratory tract infection, mcg micrograms, SAMA: short acting muscarinic antagonist, LAMA: long acting muscarinic antagonist, LTRA: leukotriene receptor antagonist, A&E Accident and Emergency

Table S2 Measures of prognostic performance algorithm in derivation and validation (hospitalisation only) for risk score thresholds of top 1%, 2%, 5%, 7% and 10% discrimination

	Derivation (N=58,619) ROC = 0.72 (95% CI 0.71, 0.74)					Validation (N=174,240) ROC=0.70 (95% CI 0.69, 0.72)				
	Risk score threshold	PPV (%)	NPV (%)	Sensitivity (%)	Specificity (%)	Risk score threshold	PPV (%)	NPV (%)	Sensitivity (%)	Specificity (%)
Top 1%	.104	19.3	98.5	11.7	99.2	.073	4.3	99.2	5.3	99.0
Top 2%	.067	13.6	98.6	16.5	98.2	.056	3.7	99.2	8.9	98.1
Top 5%	.041	8.29	98.7	25.1	95.3	.038	3.5	99.3	21.0	95.1
Top 7%	.035	6.99	98.7	29.6	93.4	.031	3.3	99.4	28.2	93.2
Top 10%	.028	5.89	98.8	35.6	90.4	.025	2.9	99.4	35.7	90.2

PPV: positive predictive value, NPV: negative predictive value

Table S3 Number of events by risk strata in derivation and validation (hospitalisation only) cohorts – with risk score thresholds of top 1%, 2%, 5%, 7% and 10%

	Derivation (N=58,619)			Validation (SAIL N = 174,240)		
	Risk score threshold	Number of people (%)	Number of events (%)	Risk score threshold	Number of people (%)	Number of events (%)
Top 1%	.104	587 (1%)	113 (19.3%)	.073	1751 (1%)	76 (4.34%)
Top 2%	.067	1173 (2%)	160 (13.6%)	.056	3,494 (2%)	128 (3.66%)
Top 5%	.041	2931 (5%)	243 (8.3%)	.038	8,727 (5%)	301 (3.45%)
Top 7%	.035	4106 (7%)	287 (7%)	.031	12,225 (7%)	405 (3.31%)
Top 10%	.028	5862 (10%)	345 (5.9%)	.025	17,427 (10%)	512 (2.94%)
Everyone	NA	58619	969 (1.65%)	NA	174240	1,434 (0.82%)