USE CASE Diagnose-Stratify-Treat: How we designed and deployed predictive AI in a clinical pathway for COPD

CHRIS CARLIN

Consultant Physician & NRS Senior Investigator, Respiratory Innovation

NHS Greater Glasgow & Clyde

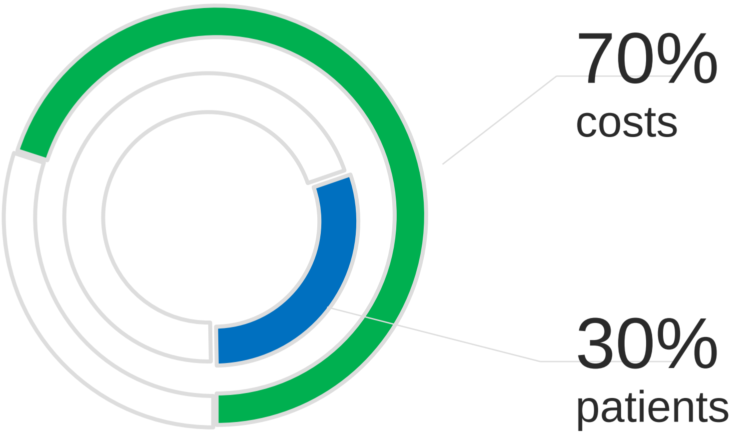
Designing and deploying predictive AI in a COPD pathway

Chris Carlin
Professor, Respiratory Innovation

NHS West of Scotland Innovation Hub
Lenus

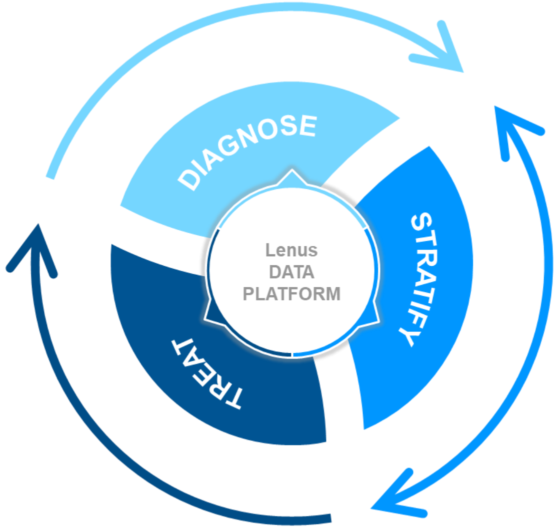
The challenge

Chronic conditions like COPD account for a disproportionate share of healthcare resources and are tied to inequalities

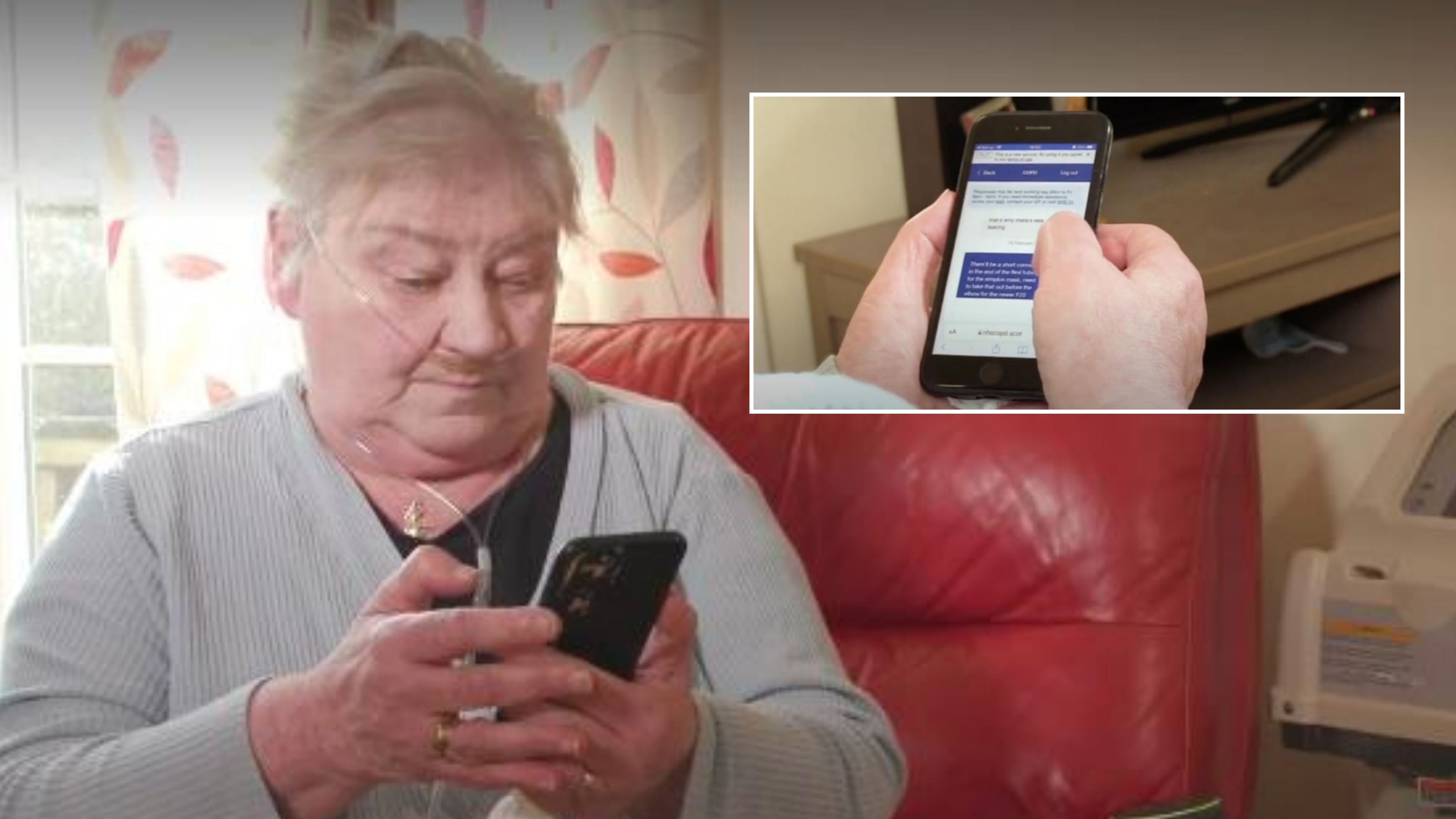


The approach

Build an end-to-end pathway that enables machine learning to support case finding and risk stratification




TEST BEDS FOR INNOVATION
Bringing together industry, academia and healthcare services to improve NHS and Social Care in Scotland
GLASGOWSafe Haven
SECURE NHS DATA RESEARCH

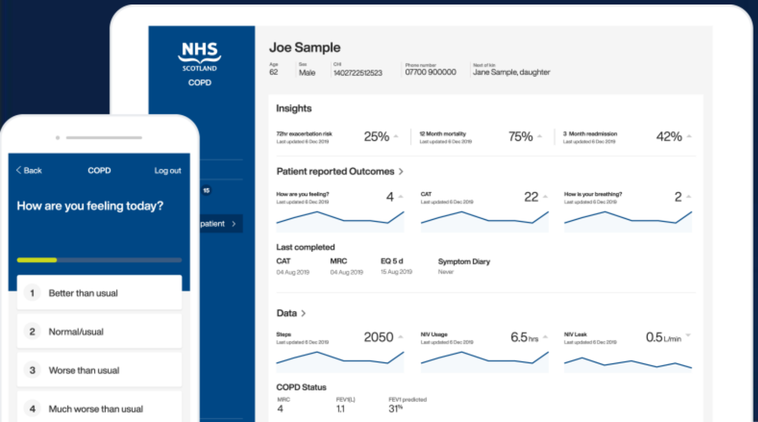

Back
COPD
Log out
nhscopd.scot

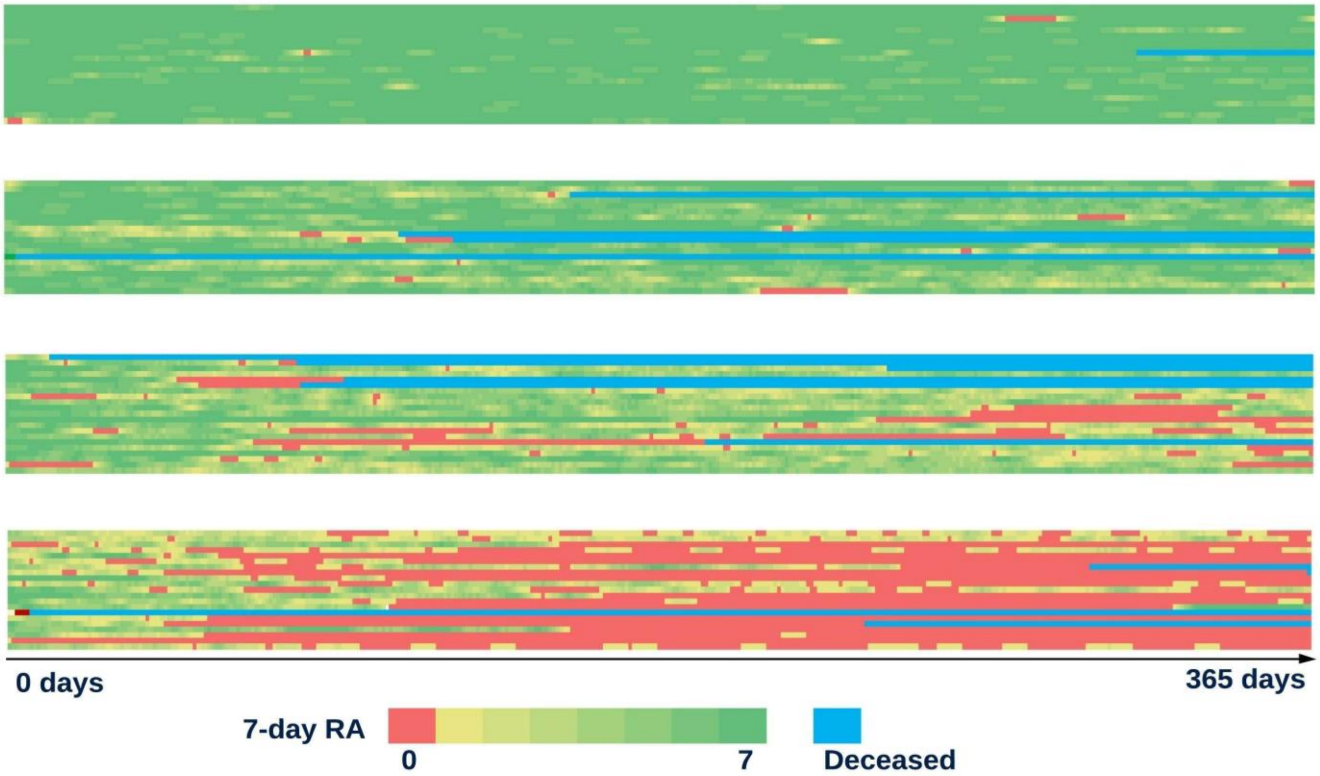
NHS
Greater Glasgow
and Clyde

Lenus

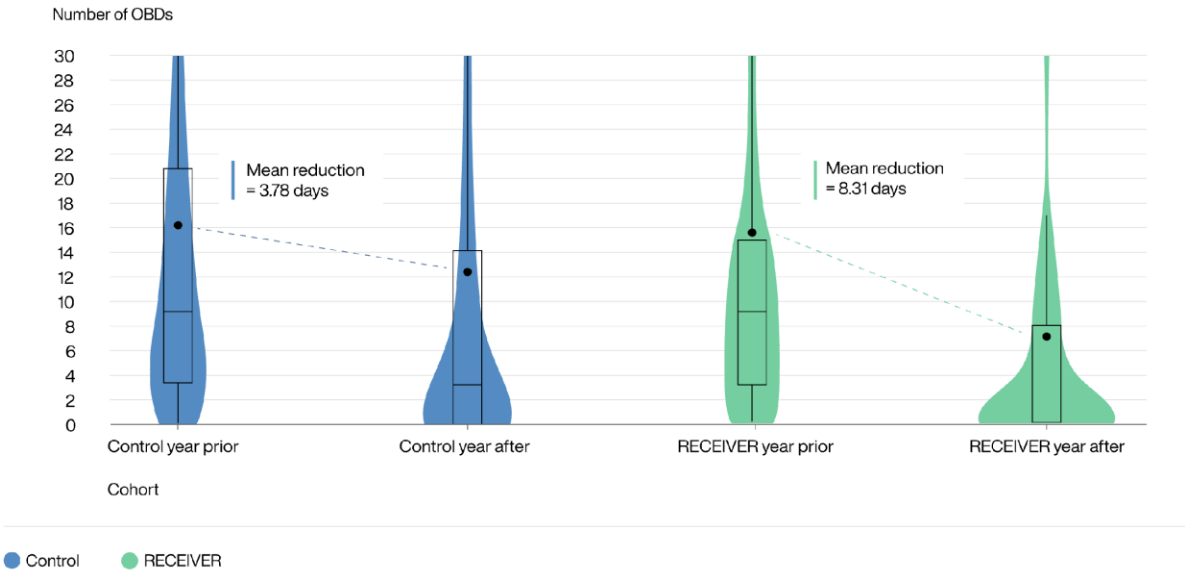
1 per week

1 per year





lenushealth.com/evidence

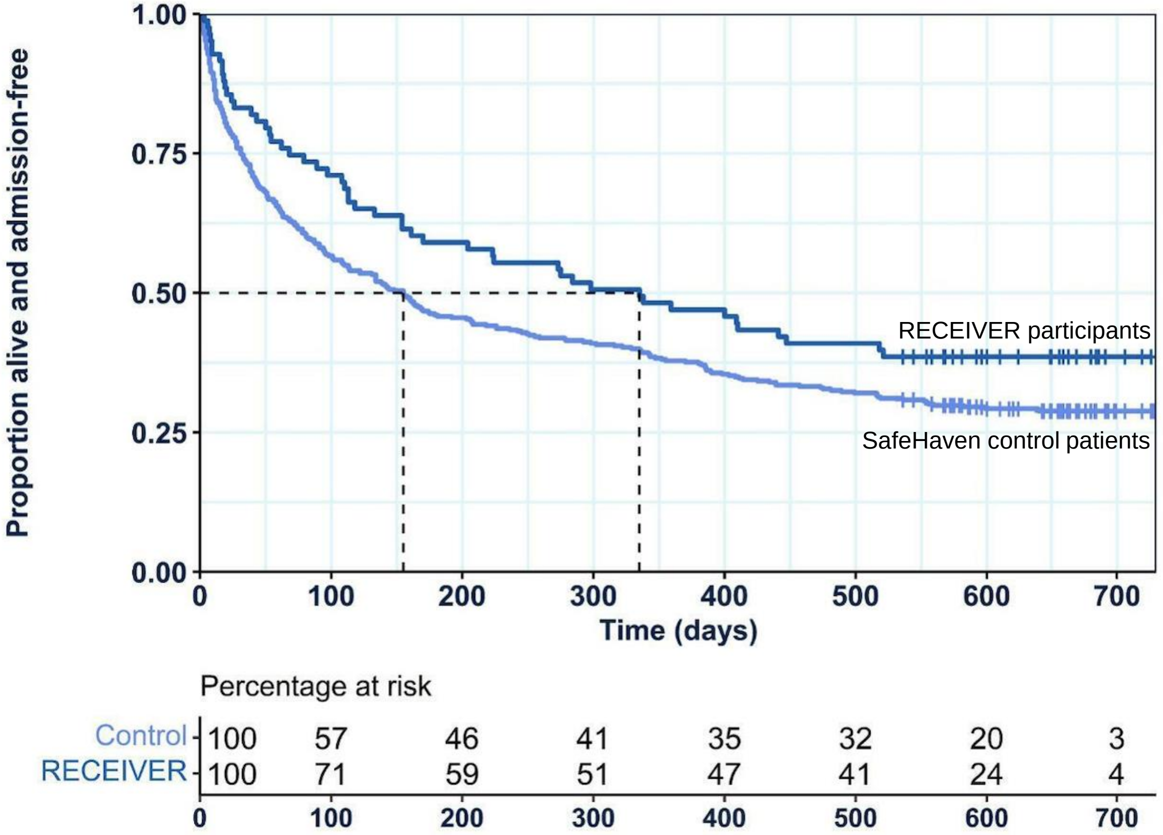


500 patients

Healthcare resource utilisation saving

£ 3.4m per year

lenushealth.com/evidence



lenushealth.com/evidence


"This is just the tip of the iceberg…"

Predicting 12-month mortality in a Scottish COPD cohort

Shane Burns[1], Grace Cox[1], David J Lowe[2], Anna Taylor[2], Paul McGinness[1], Chris Carlin[2]
[1] StormID, Edinburgh [2] Respiratory & Emergency Medicine, NHS Greater Glasgow & Clyde

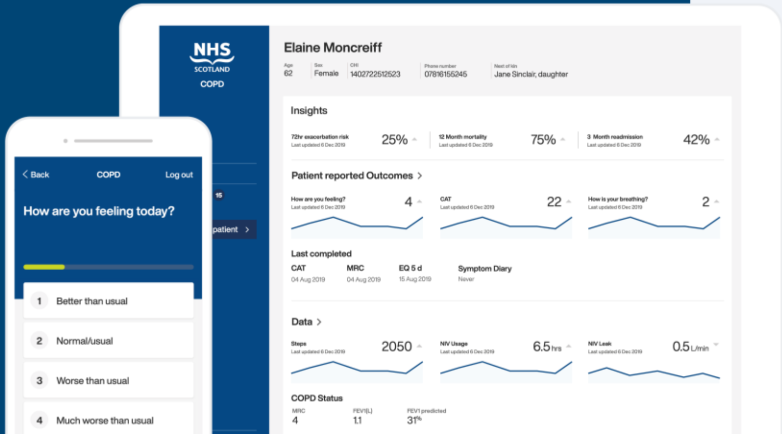
De-identified cohort established from NHS GG&C SafeHaven

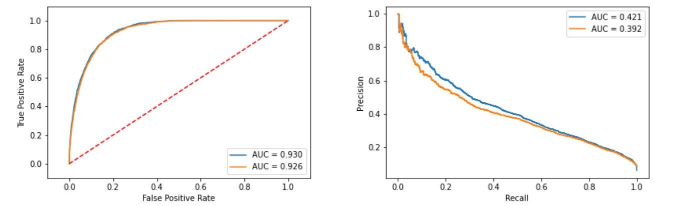
- NHS GG&C: largest healthcare organisation in UK; 1.2m population, high prevalence of COPD with high admission & mortality rates[1].
- 55531 patients with COPD (ICD-10 J44* in NSS SMR01 dataset)
- Index severe exacerbation + minimum 12 months follow-up data
- Demographic, coded diagnoses, hospital admission, prescribing and laboratory data.

Predictive model co-designed by clinician and data science team

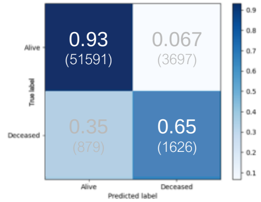
- Binary target variable: alive or deceased at 12 months following severe exacerbation?
- Class stratified15% of data (7884 patients) = hold out test dataset.
- Remaining 85% of data used for model training and k-fold cross validation.
- ML models applied, with XGBoost demonstrating best performance.
- Model features and performance on hold out test dataset presented.

1. *Scottish Atlas of Healthcare Variation, ISD Scotland. https://bit.ly/COPD_ScotAOV*

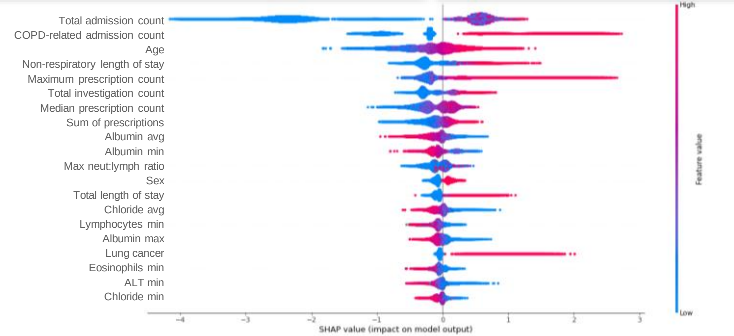




Receiver operator characteristic and precision recall curves Red line on ROC curve is a no skill model and blue lines on both curves are baseline model performance on the holdout dataset (reported as the area under curve, AUC). Prediction of alive vs deceased at 12 months following a severe COPD exacerbation exceeds published comparators, with high accuracy and precision supporting clinical implementation. Model performance is retained in dropout analysis (orange lines) when comorbidity feature set is rationalised from all coded diagnoses to a data-driven 'top 20' diagnosis list which can be realistically captured in the LenusCOPD clinical user interface



Model confusion matrix from patient-year observations Actionable AI insights for an individual can be derived. For example, a patient predicted by this model to be alive has a post-exacerbation 1.7% 12-month mortality. A patient predicted to be deceased has a post-exacerbation 30.5% 12-month mortality, which triggers MDT discussion and prioritisation for anticipatory care planning. 9% of patients fall into this deceased predication category, which is a manageable MDT workload.



Shapely values (SHAP) feature importance plot. Global model interpretability is shown as the impact of bio-plausible model features on the overall 12-month mortality prediction model. Each dot is a single data point in the train data (blue = low values, red = high values) and increasing distance from the zero SHAP importance line denotes stronger influence on the model. Feature importance utility includes prioritising electronic health record integrations and clinician-inputted source data requirements for live model risk scores. Model output also includes individual patient SHAP plots, providing local interpretability and decision support to the COPD MDT.

Our AI-based COPD 12-month mortality prediction model demonstrates excellent performance: integration of model scores and explainability plots within LenusCOPD "AI Insights" dashboard extension for MDT use is planned.

Experience with MDT presentation of model risk scores in our proposed 'DYNAMIC-AI' implementation-effectiveness clinical investigation will inform adoption and further evaluations of AI insight-based decision support.

Additional models - 3-month admission and 72-hour exacerbation risk - are in advanced development.

Daily patient-reported outcome with wearable and respiratory therapy monitoring data from scale-up of the LenusCOPD service (support.nhscopd.scot) will provide reference ground truth event data and input features for continuous model improvements.

Predicting 3-month respiratory readmission in a Scottish COPD cohort

Shane Burns, Anna Taylor*, Grace Subašić, Paul McGinness, David J Lowe*, Chris Carlin*
LenusHealth, Edinburgh and *Respiratory & Emergency Medicine, NHS Greater Glasgow & Clyde
Get in touch lenushealth.com or christopher.carlin@ggc.scot.nhs.uk

Lenus

NHS SCOTLAND

De-identified cohort established from NHS GG&C SafeHaven

NHS GG&C Largest healthcare organisation in UK, serving 1.2m population. High prevalence of COPD, with high admission & mortality rates.

Dataset Demographics, coded diagnoses, hospital admissions, prescribing and laboratory data from 33148 patients with COPD who had 120252 respiratory-related admissions.

Predictive AI model co-designed by clinician and data science team

Binary target variable Respiratory-related readmission at 3 months post discharge.

Model training 85% of data used, with k-fold cross validation.

Holdout test dataset Class stratified 15% of data to evaluate models' performance and fairness.

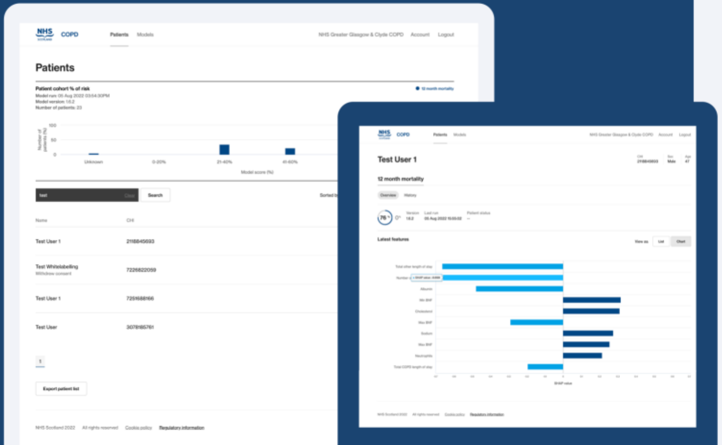
XGBoost model Selected as demonstrated best performance.

Conclusions

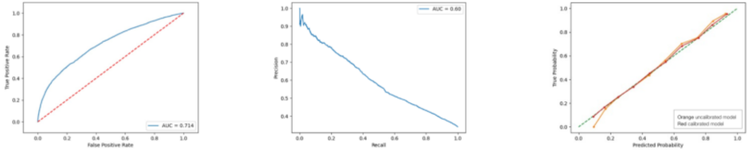
Model's performance and utility Ready for adoption within COPD MDTs.

Operationalising live AI models Co-designed cloud-based COPD AI insights app ready to be deployed.

DYNAMIC-AI Patient acceptability and technical feasibility will be determined alongside a range of safety and utility secondary objectives in this implementation-effectiveness trial.



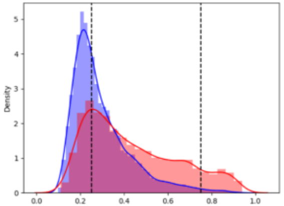
Receiver operator characteristic, precision recall and calibration curves from holdout test dataset



Distribution of model inference probabilities

Blue no readmission **Red** readmission

Selecting probability threshold of 0.25 tunes model to bring forward most patients with readmission, with reduced specificity.
Selecting probability threshold of 0.75 tunes model to high specificity, but misses large n of patients who have a readmission.
Clinicians & data scientists can collaborate to adapt model outputs depending on use case and service capacity.

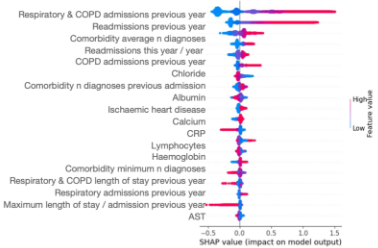


Probability threshold	Number of patients predicted 3-month readmission / 100 patients		
	Correct	Incorrect	Missed
0.25	27	34	7
0.5	12	5	23
0.75	4	0	31

Global model explainability data

SHAP plot of model feature importance

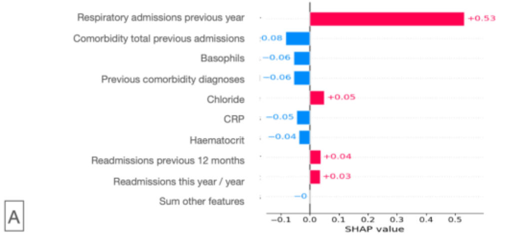
Feature review provides actionable insights: respiratory failure, requirement for respiratory support and anxiety are highlighted as diagnoses associated with high-risk for readmission.

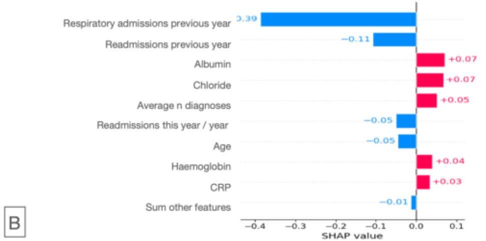


Individual patient explainability data

SHAP plots of model feature importance

A Correct prediction readmission **B** Correct prediction no readmission





Model development dataset

- NHS Greater Glasgow & Clyde Safe Haven DYNAMIC cohort.
- All patients with COPD resident NHS GG&C 1st Jan 2013 - 31st Dec 2019.
- Datasets: demographics, prescribing, laboratory, hospital admission and comorbidity data.

Changes from previous model:
- None

Model formulation

- Binary classification problem predicting 12-month mortality (Class 0 = alive at 12 months; Class 1 = deceased at 12 months).
- Model inference per patient per model run yields a score ranging from 0-1, which can be interpreted as the probability of patient being deceased at 12 months. This probability is converted to a percentage in the COPD AI Insights App.
- XGBoost algorithm (decision tree based) with hyperparameters tuned to the clinical problem.
- All model parameters and metrics are logged in the model tracking system.

Changes from previous model: None

Model training and validation

85% of dataset (~ 39K patients)

Model evaluation

15% of dataset – 'holdout' test (~ 7K patients)

- In order to prevent data leakage, the full dataset was split such that an individual patient may only appear in either the train or holdout test cohort.
- Demographic make-up (mean age = 68 yr, 57% female) and class balance (proportion of patients alive vs deceased at 12 months, 7%) matched for training and holdout test datasets.
- In order to prevent data leakage, model validation during training was performed using K-fold cross validation, in which the train dataset is split patient-wise into K folds.
- Model performance and associated metrics for model approval and presentation in the COPD AI Insights App are reported from the holdout test cohort.

Changes to this approach from previous model: None

Feature engineering

- Features are a combination of one-hot encoding, target encoding and domain-driven feature engineering techniques. See Table 5 for the model features and applied engineering techniques
- Where target encoding or other data aggregations are performed across patients, this uses a K-fold approach to prevent data leakage. Aggregations from the training data were applied to holdout test data and will be applied to inference data.
- Data quality issues and missingness were explored and remedied where necessary.
  - A maximum missingness threshold of 40% was applied to lab data to determine whether a lab test feature could be included in the model.
  - A small amount of poor-quality pharmacy data was removed.
  - Missing data for the selected laboratory-derived features (those with missingness less than 40%) was allowed during model training. This was the case as the chosen algorithm was sparsity aware, meaning the missing data did not need to be imputed.
  - Missing data in all other features was infilled appropriately. For all other features, missing data were replaced with zero as the context for missingness in those cases implied zero occurrence.
  - Patient age was scaled using the training data. The holdout test dataset was infilled and scaled using properties of the training data to prevent data leakage.
  - Multiple admissions corresponding to the same hospital admission event were consolidated as one admission. This occurs, for example, when a patient is transferred between different hospital facilities as part of the same admission.
- Feature interaction constraints are imposed on the 'days since' lab test features. Each 'days since' feature can only interact with the corresponding lab test, the patient's age, and the patient's sex.
- Hospital admission diagnosis codes are collated to align to comorbidities as recorded in Lenus COPD service.

Changes from previous model: None

Model performance metrics

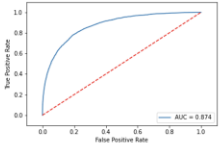


Figure 3 - Receiver Operating Characteristic (ROC) curve of the candidate model, evaluated on the holdout test cohort.

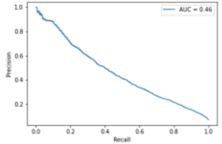


Figure 4 - Precision-Recall (PR) curve of the candidate model, evaluated on the holdout test cohort.

Probability threshold	Class	Precision	Recall	F1 Score	ROC-AUC	PR-AUC	Brier Loss
0.25	Deceased	0.46	0.44	0.45	0.87	0.46	0.048
	Alive	0.96	0.96	0.96			
0.5	Deceased	0.70	0.2	0.32	0.87	0.46	0.048
	Alive	0.99	0.94	0.97			
0.6	Deceased	0.78	0.16	0.26	0.87	0.46	0.048
	Alive	1	0.94	0.97			
0.8	Deceased	0.88	0.08	0.15	0.87	0.46	0.048
	Alive	1	0.94	0.97			

Table 2 - Candidate model performance with different probability thresholds, evaluated on the holdout test cohort.

Comments:
- The candidate model performance is in line with expectations.
- Candidate model performance is the same as the current model.

Model calibration, performance and thresholding

- If a model is perfectly calibrated, the inference probabilities will match the true probability of the event, e.g., half of patients with a model score of 0.5 will truly be deceased within 12 months.
- The candidate model was calibrated during cross validation to maximise matching between inference probability and true probability.
- Thresholding is the process of converting probability outputs to class labels by choosing a cut off value (values less than this output as class 0, values greater than this as classs1). The choice of threshold will affect the model performance metrics and expected clinical workload
- Several model performance metrics are presented in this report. Of particular importance for this clinical use-case are:
  - The area under the precision recall curve (PR-AUC). Other metrics tend to be inflated and misleading when class imbalance is high.
  - Expected numbers of patients brought forward correctly/incorrectly and missed for different threshold levels. This is important as it simulates the expected clinician workload.

Changes to this approach from previous model:
- Updated calibration method to use custom patient folds versus the built in sklearn record-wise fold generation. This gave a slight improvement to model calibration.

Model explainability and fairness

- In accordance with ethical AI principles, model explainability and fairness were evaluated
- Global model explainability describes which model features are important to the overall model across the entire training cohort.
- Local model explainability describes what is important on an individual prediction level.
- This allows for interrogation of the model prediction and to identify potential biases.
- Model fairness relates to performance on different sub-groups of interest within the population.
- These sub-groups may relate to demographics (e.g., sex, age, ethnicity) or clinical factors such as the presence/absence of certain comorbidities.
- The model can be re-trained if potential inequalities are identified, tuning the loss functions to ensure parity between groups.

Changes to this approach from previous model:
- The shap values for both global and local explainability are now calculated by averaging the shap values from each of the base models that form the final calibrated model.

Model QA (technical)

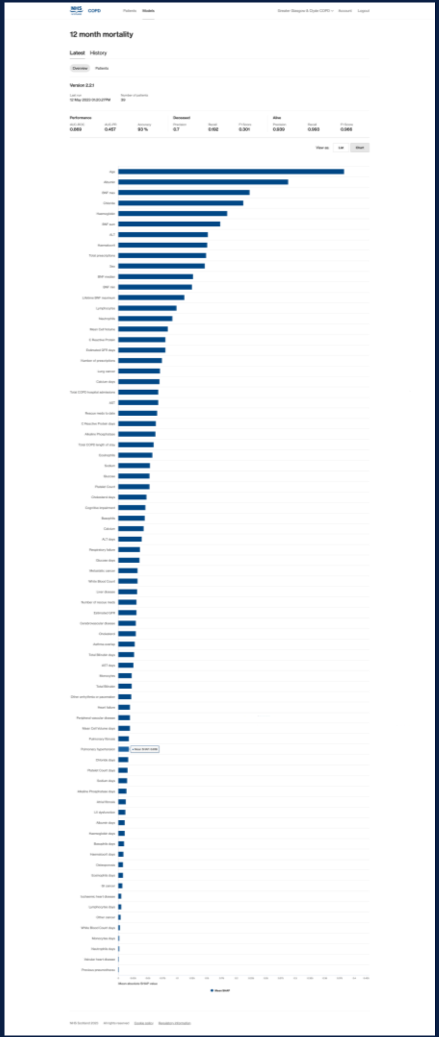
The model was reviewed internally by the Lenus Engineering team on 31/10/2022 and deemed an acceptable candidate for clinician model approval. The detailed report of the technical review is available.

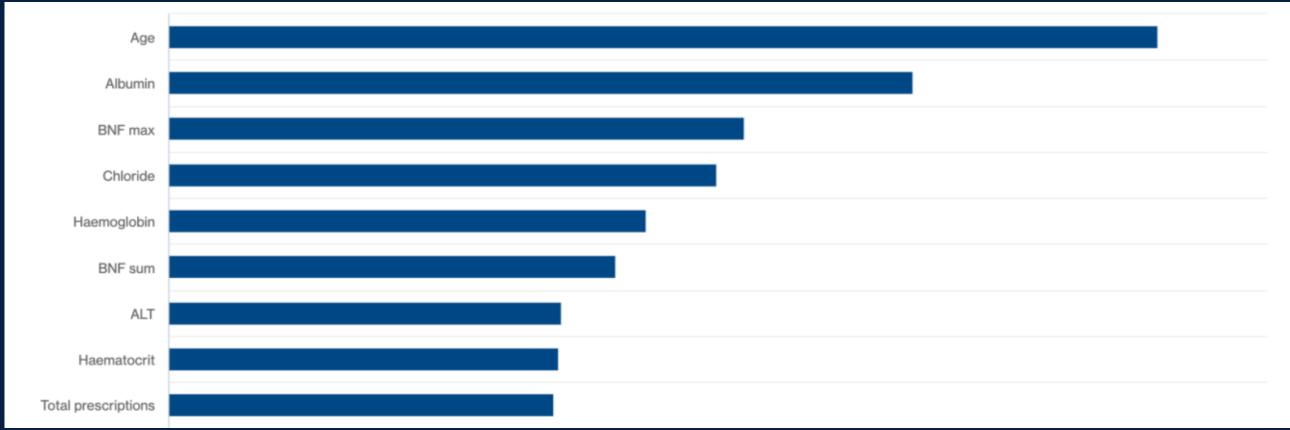
Model approval review

Highlights and discussion points

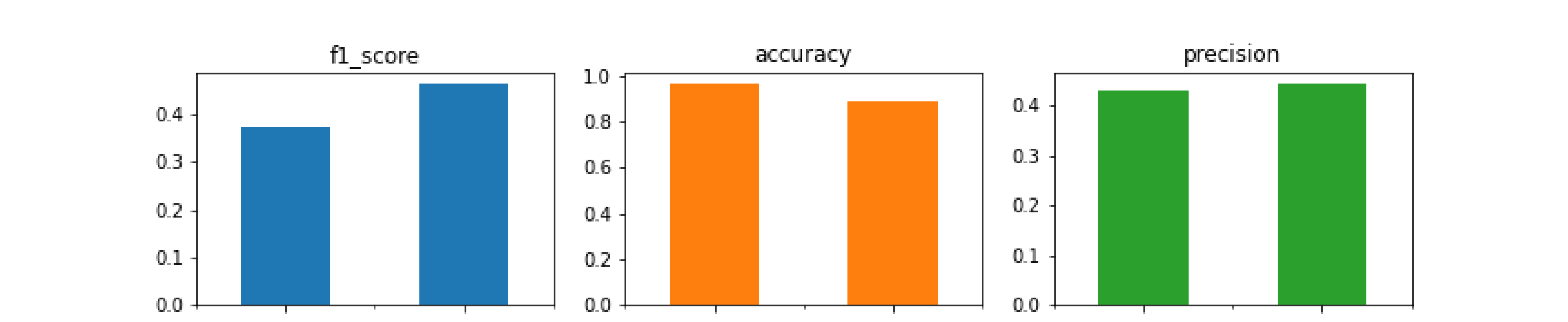
- The candidate model is well calibrated. This can be seen visually from the calibration curve, and from the low Brier Loss.
- There is reasonable separation of inference probabilities for the two classes.
- A clinical decision based on available resources is required for the most appropriate threshold.
- Model performance is in line with expectations and unchanged from the current model.
- Checks in cross-validation indicate that no data leakage or overfitting to training data has occurred.
- The candidate model global and local explainability appears bio-plausible. To be verified with clinicians.
- The candidate model performs slightly better on men versus women.
- The candidate model performs better on the over 65s age group verses under 65s.
- The candidate model performs slightly better on less deprived SIMD groups.
- The context of having had a specific lab test can be more important than the value itself. There are more missing values for people who did survive the following year implying that people who are generally sicker, get tested for more things. This motivated the decision to impose interaction constraints on the lab 'days since' variables to help improve model generalisability. In a different healthcare setting it may be more routine to take certain labs test measurements so using the 'days since' feature without interaction constraints may introduce a systematic bias.

Model explainability





Model fairness



Dynamic AI study

We're running a study to find out if computer-based problem solving, or artificial intelligence (AI), can help us care for patients with COPD.

About the study

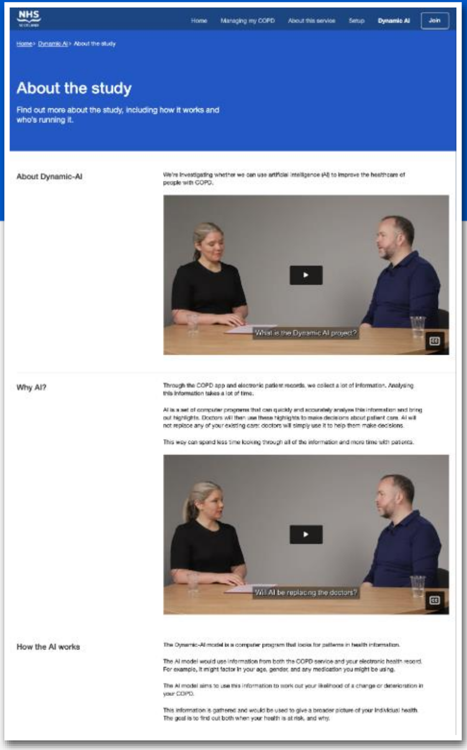
Find out more about the study, including how it works and who's running it.

Join the Dynamic-AI study

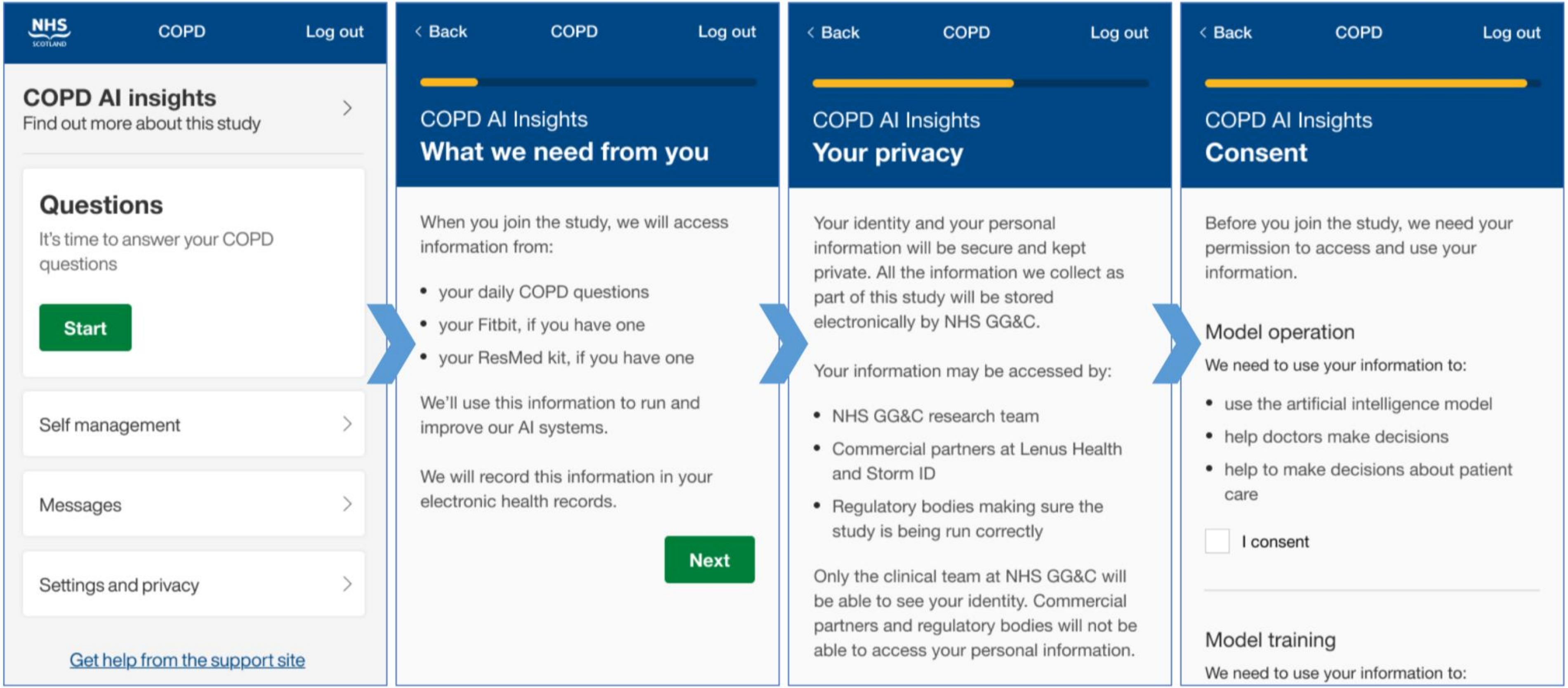
Dynamic-AI is available by invitation only for people with COPD in Greater Glasgow and Clyde.

Your privacy

How we're keeping your information safe, secure, and confidential.



DYNAMIC-AI
Support website



NHS SCOTLAND
COPD
Log out
COPD AI insights
Find out more about this study
Questions
It's time to answer your COPD questions
Start
Self management
Messages
Settings and privacy
Get help from the support site
Back
COPD
Log out
COPD AI Insights
What we need from you
When you join the study, we will access information from:
• your daily COPD questions
• your Fitbit, if you have one
• your ResMed kit, if you have one
We'll use this information to run and improve our AI systems.
We will record this information in your electronic health records.
Next
Back
COPD
Log out
COPD AI Insights
Your privacy
Your identity and your personal information will be secure and kept private. All the information we collect as part of this study will be stored electronically by NHS GG&C.
Your information may be accessed by:
• NHS GG&C research team
• Commercial partners at Lenus Health and Storm ID
• Regulatory bodies making sure the study is being run correctly
Only the clinical team at NHS GG&C will be able to see your identity. Commercial partners and regulatory bodies will not be able to access your personal information.
Back
COPD
Log out
COPD AI Insights
Consent
Before you join the study, we need your permission to access and use your information.
Model operation
We need to use your information to:
• use the artificial intelligence model
• help doctors make decisions
• help to make decisions about patient care
I consent
Model training
We need to use your information to:

COPD Insights

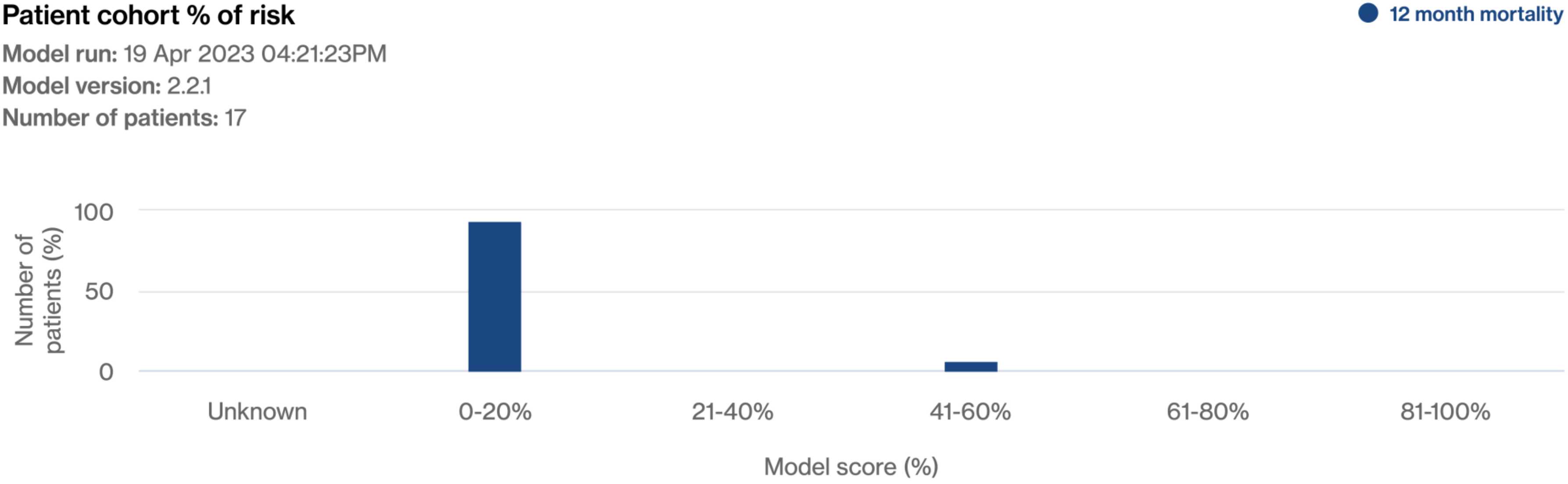
Sign in to the COPD Insights service.

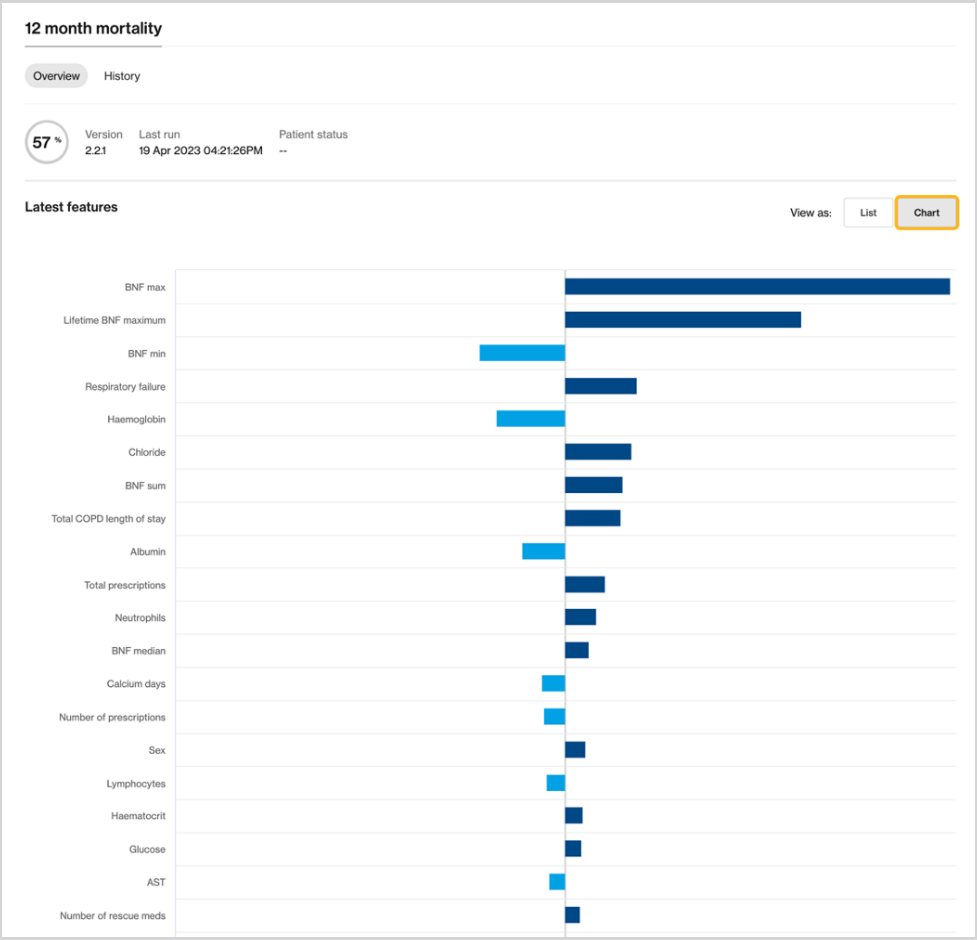
Log In

Patients

Patient cohort % of risk

Model run: 19 Apr 2023 04:21:23PM
Model version: 2.2.1
Number of patients: 17

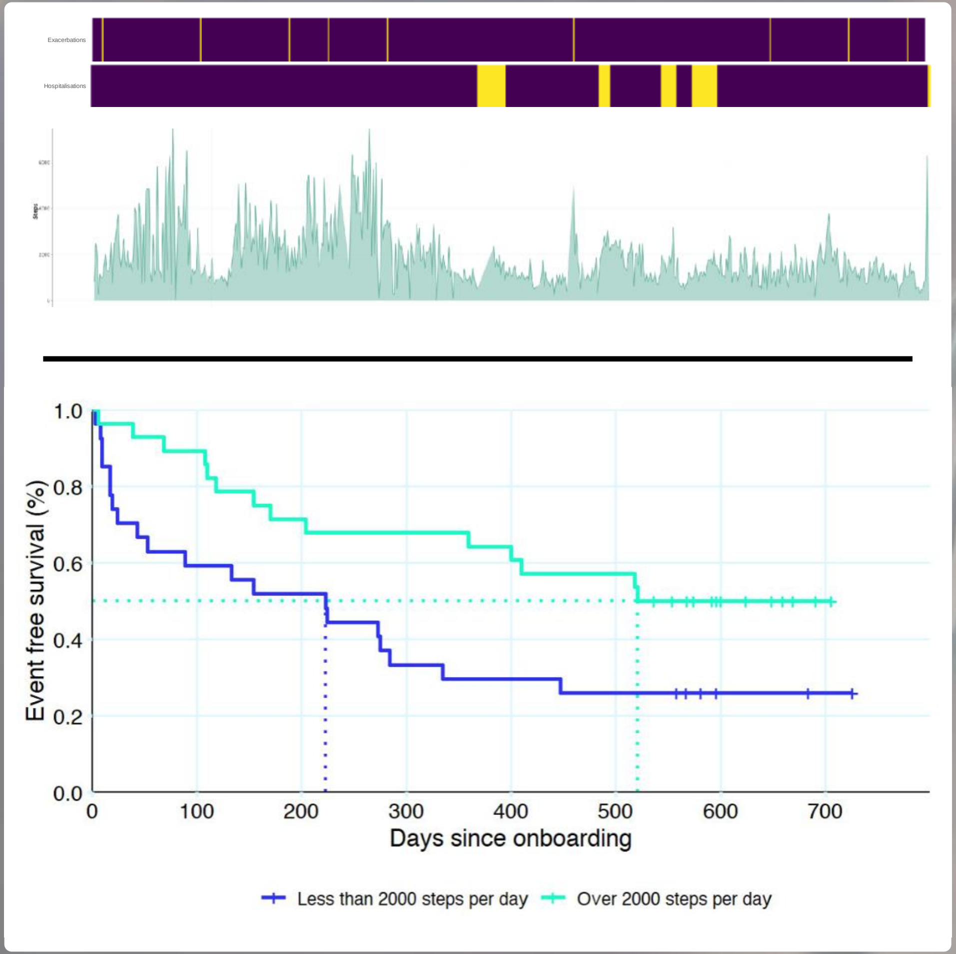




Hi Mr
Thanks for joining the DYNAMIC-AI trial. We got the first run of data through from it, and it's really interesting, with nothing worrying. It did flag one thing - that it might be worth checking your overnight breathing or blood gases again at some point. We could have a chat about that - not urgent, but I'd have clinic space + could give you a call tomorrow or thursday sometime, if any time either day would suit you? Chris

Chris Carlin - 25 April 2023 13:28


“This is just the tip of the iceberg…”



RECEIVER trial, unpublished data

12-month admission vs no admission

Parameter	Mean- No Admissions	Mean- Had admissions	SD- No admissions	SD- Had admission	p value (Mann-Whitney test)	Hodges Lehmann estimate	Hodges Lehmann estimate/ Mean- No admissions
IE ratio (Max)	64.189	58.150	19.809	16.224	**6.71e-37**	5.9999795	0.0934736
IE ratio (Median)	40.133	34.819	10.053	7.015	**1.04e-111**	5.9999555	0.1495018
IE ratio (95th	54.056	45.064	14.158	10.525	**1.55e-154**	9.0000309	0.1664946
Minute ventilation (Median)	9.341	8.876	2.106	2.243	**3.84e-53**	1.1250278	0.1204398
Minute ventilation (95th)	12.107	11.517	3.277	3.881	**7.73e-36**	1.2500347	0.1032489
Apnea-hypopnea index	1.855	1.114	3.242	2.898	**1.14e-109**	0.5000382	0.2695623
Hypopnea index	1.505	0.606	2.948	1.263	**5.80e-129**	0.4000876	0.2658390
Patient triggered expiration (%)	67.994	84.763	22.033	20.435	**2.26e-272**	-16.4999548	-0.2426678
Patient triggered inspiration (%)	76.343	81.758	21.375	24.747	**7.00e-67**	-5.0000172	-0.0654941
EPAP value (Median)	7.669	9.019	1.734	2.919	**8.22e-50**	-1.1999474	-0.1564673

90-day admission vs no admission

Parameter	Mean- 90 days pre-admission	Mean- Other	SD- 90 days pre-admission	SD- Other	p value (Mann-Whitney test)	Hodges Lehmann estimate	Hodges Lehmann estimate/ Mean- 90 day pre-admission
Leak value (Max)	0.927	0.622	0.693	0.675	**5.37e-27**	0.2800298	0.3020818
Leak value (95th)	0.460	0.244	0.479	0.378	**1.12e-31**	0.1399548	0.3042496
Minute ventilation (95th)	13.358	10.919	4.999	3.222	**2.43e-20**	1.7500286	0.1310098
Apnea-hypopnea index	2.866	0.545	5.049	1.236	**1.69e-46**	0.3999887	0.1395634
Apnea index	1.457	0.178	3.350	0.722	**3.02e-48**	0.0999804	0.0686207
Hypopnea index	1.384	0.353	2.041	0.706	**1.80e-44**	0.2000183	0.1445219
Respiratory rate (Max)	30.129	25.042	6.067	4.523	**4.57e-89**	5.6000694	0.1858697
Respiratory rate (Median)	19.436	16.695	4.875	2.949	**6.00e-34**	2.2000544	0.1131948
Respiratory rate (95th)	25.253	20.555	6.166	4.243	**3.32e-74**	4.7999485	0.1900744
Minutes spO2 below 88 percent	2.502	1.994	1.026	0.850	**1.84e-33**	0.4805169	0.1920531
Seconds spO2 below dynamic threshold	0.056	0.050	0.020	0.021	**6.83e-13**	0.0060204	0.1075064
EPAP value (Median)	9.803	8.764	2.691	2.946	**3.17e-24**	0.9599499	0.0979241

7-day admission vs no admission

Parameter	Mean- week pre-admission	Mean- Other	SD- week pre-admission	SD- Other	p value (Mann-Whitney test)	Hodges Lehmann estimate	Hodges Lehmann estimate/ Mean- week pre-admission
Respiratory rate (Max)	26.189	29.550	5.351	6.225	**2.81e-07**	-3.399929	-0.1150568
Respiratory rate (95th)	21.618	24.583	5.128	6.421	**7.98e-06**	-2.999963	-0.1220340

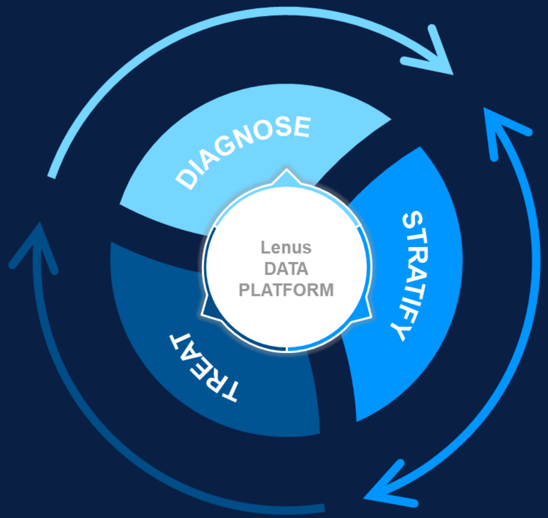
RECEIVER trial, unpublished data

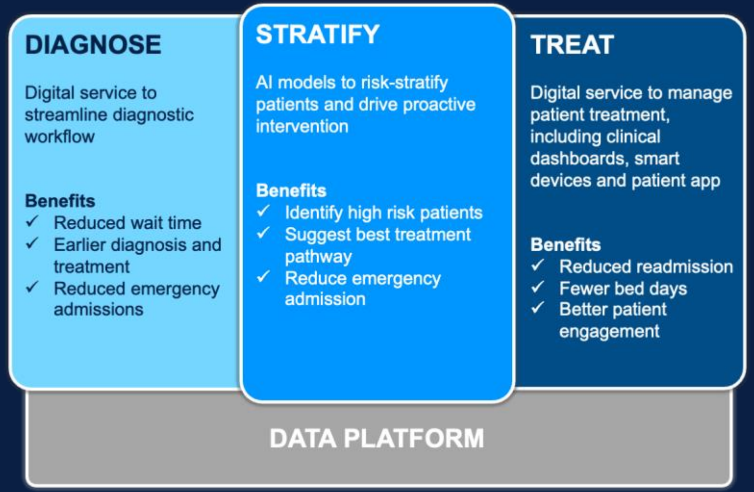
POLARIS

Prevention of **CO**PD **L**ung **A**ttacks by Ea**R**ly **I**ntervention **S**trategies

Predictive AI in an end-end COPD pathway

Exemplar for long term condition care





DYNAMIC-AI trial

Co-design live AI insights at point of care in COPD MDT

Clinical investigation acceptability, feasibility and utility

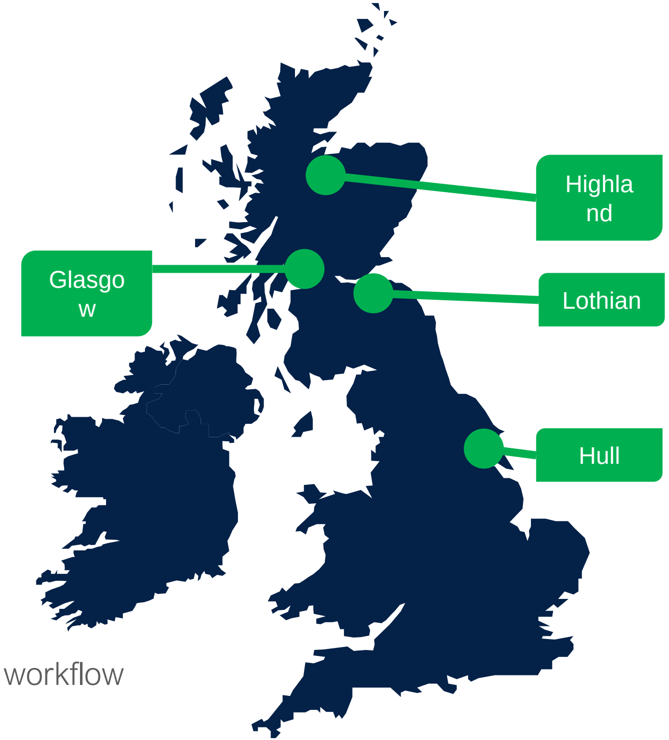
End-end pathway

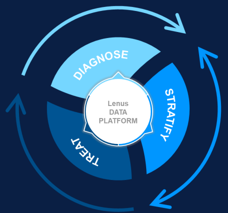
LenusCOPD service sustained use, improved outcomes

Scale-up recover and re-orientate diagnostics, other sites

Test bed

Implement and validate AI models and other solutions within clinical workflow





Designing and deploying predictive AI in a COPD pathway

Chris Carlin
Professor, Respiratory Innovation

NHS Scotland West of Scotland Innovation Hub

Lenus

INTELLIGENT HEALTH UK 2023

Breaking down the barriers between tech and healthcare