Online Supplementary Material

Online supplementary table 1: DECAF indices as predictors of inhospital mortality

Index	B	OR (95%CI)	P value	Score
Dyspnoea				
eMRCD 1-4		1	<0.0001	0
eMRCD 5a	1.13	3.10 (1.78- 5.40)	<0.0001	1
eMRCD 5b	2.17	8.79 (5.13-15.03)	<0.0001	2
Eosinopenia (<0.05 x10^9/L)	0.90	2.45 (1.61-3.72)	<0.0001	1
Consolidation	0.90	2.45 (1.66-3.62)	<0.0001	1
Acidaemia (pH <7.3)	1.35	3.87 (2.38-6.31)	<0.0001	1
Atrial Fibrillation	0.85	2.35 (1.53-3.60)	<0.0001	1

Online supplementary table 2: Sensitivity and specificity of DECAF in pneumonic exacerbations of COPD.

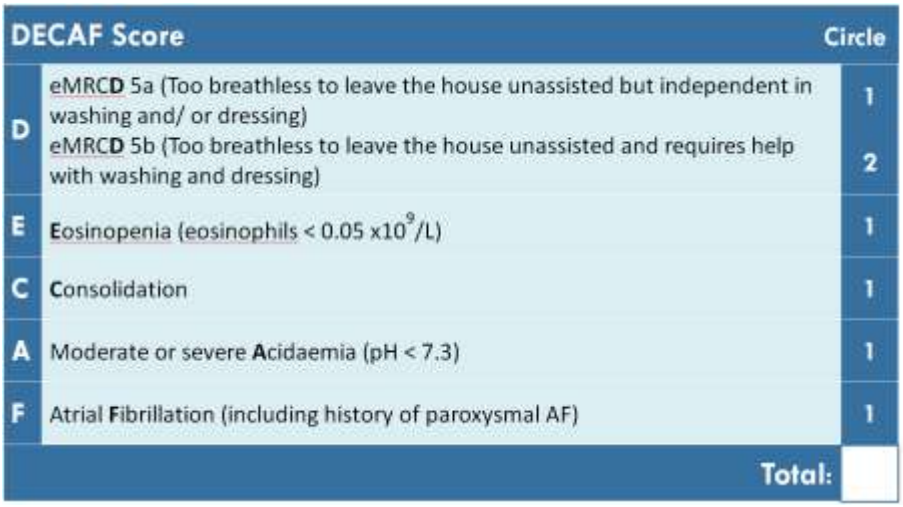
	Validation cohorts		Derivation and validation cohorts	
DECAF score	Sensitivity	1 - Specificity	Sensitivity	1 - Specificity
0	N/A	N/A	N/A	N/A
1	1.00	0.84	0.98	0.82
2	0.87	0.48	0.82	0.44
3	0.55	0.20	0.54	0.17
4	0.20	0.052	0.21	0.043
5	0.014	0.0024	0.0078	0.0015
6	0	0	N/A	N/A

Online supplementary table 3: Inpatient and 30-day mortality by DECAF and CURB-65 score in those with pAECOPD, derivation and validation cohorts.

	Inpatient death, n (%)		30-day death, n (%)	
Score	DECAF	CURB-65	DECAF	CURB-65
0	N/A*	3/55 (5.5)	N/A*	4/55 (7.3)
1	2/122 (1.6)	14/182 (7.7)	4/122 (3.3)	20/182 (11.0)
2	21/267 (7.9)	31/264 (11.7)	25/267 (9.4)	39/264 (14.8)
3	36/215 (16.7)	50/219 (22.8)	42/215 (19.5)	57/219 (26.0)
4	43/129 (33.3)	24/57 (42.1)	48/129 (37.2)	24/57 (42.1)
5	26/53 (49.1)	7/11 (63.3)	29/53 (54.7)	5/11 (45.5)
6	1/2 (50)	N/A	1/2 (50)	N/A
Total	129/788 (16.3)		149/ 744 (18.9)	

*The lowest possible DECAF score in those with pAECOPD is 1.

Online supplementary figure 1: The DECAF score



DECAF Score		Circle
D	eMRC**D** 5a (Too breathless to leave the house unassisted but independent in washing and/ or dressing)	1
	eMRC**D** 5b (Too breathless to leave the house unassisted and requires help with washing and dressing)	2
E	**E**osinopenia (eosinophils < 0.05 x10^9/L)	1
C	**C**onsolidation	1
A	Moderate or severe **A**cidaemia (pH < 7.3)	1
F	Atrial **F**ibrillation (including history of paroxysmal AF)	1
	Total:	

CHARMS CHECKLIST

Source of Data

1) Source of data (e.g. cohort, case-control, randomised trial participants, or registry data)

- The external validation cohort was prospective, and individually adequately powered. The internal validation cohort was partially prospective, with retrospective extension.

Participants

1) Participant eligibility and recruitment method (consecutive participants, location, number of centres, setting, inclusion and exclusion criteria)

Eligible patients analysis

- All eligible patients were included in the validation cohorts, except those that did not have complete data for all DECAF indices. This was approximately 1% of the population, and mainly comprised of patients with SpO2 92% or less in whom arterial blood gas analysis was not performed

Eligible patients excluded

- Exclusion criteria were few. For the internal validation study, patients were excluded for the following reasons: survival <1 year n=27 (12 lung cancer, 3 end stage dementia, 3 metastatic cancer, 2 metastatic bladder cancer, 2 idiopathic pulmonary fibrosis, 1 metastatic renal cancer, 1 metastatic bower cancer, 1 metastatic rectal cancer, 1 oesophageal cancer, and 1 mesenteric cancer), less than ten pack year smoking history n=24, spirometry not obstructive= 42. Ten patients had no ABG results, but had supplemental oxygen or oxygen saturations that were too low to assume a DECAF acidaemia score of 0. One patient had no eosinophil count. Robust data is not available for the external validation cohort.

Consecutive patients

- Extensive efforts were made to capture consecutive patients, including a broad coding search. Patients were captured by daily screening (Monday to Friday) on admission units and medical wards (external validation cohort) by a dedicated team.
- In the internal validation cohort, patients were mainly identified retrospectively using a broad coding search, with cross referencing to patients identified by clinical staff who routinely review patients admitted with AECOPD. This methodology was compared to prospective screening over three months, showing superior capture overall, and only one eligible patient was identified by prospective screening alone.

Location, centres, setting, and inclusion and exclusion criteria

- Six UK centres were involved: the two sites included in the derivation study took part in the internal validation cohort, and four geographically distinct hospitals took part in the external validation. All patients in the study were recruited from secondary care. Inclusion and exclusion criteria are described.

2) Participant description

- Detailed description of participants by different sites in table 2

3) Details of treatments received, if relevant

- Medical treatment for acute exacerbations of COPD included antibiotics, steroids, nebulised bronchodilators, and non-invasive ventilation. In creating a score that is intended for use to guide management, it is not appropriate to include acute treatments as predictors. The research team did not influence clinical treatment.

4) Study dates

- Given for each site in table 2

Outcome to be predicted

1) Definition and method for measurement of outcome

- Outcome clearly defined- Inpatient death

2) Was the same outcome definition (and method for measurement) in all patients?

- Yes

3) Type of outcome single or combined endpoints?

- Single endpoint

4) Was the outcome assessed without knowledge of the candidate predictors (I.e., blinded)?

- The DECAF indices were apparent to the team reporting in-hospital death, however in-hospital death is inherently objective, therefore the risk of bias is minimal / absent.

5) Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?

- No

6) Time of outcome occurrence or summary of follow-up

- Inpatient death. Presence or absence of outcome captured in all patients.

Candidate predictors

1) Number and type of predictors (e.g. Demographics, patient history, physical examination, additional testing, disease characteristics)

- Candidate predictors refers to indices for the derivation study, not the validation study, so the number of predictors is not relevant here. The analysis of inpatient mortality was only performed with the five DECAF indices.

2) Definition and methods for measurement of candidate predictors

- Candidate predictors were described in the derivation study. In the validation study, definitions and methods of measurement are provided. Each research site was provided with a data collection guide which included definitions of terms and diseases. eMRCD score as per table 1, eosinophil count cut-off provided, presence of chest radiograph based on assessment from consultant post-take ward round, acidaemia based on arterial blood gas analysis, and atrial fibrillation based on electrocardiogram and/ or history of (paroxysmal) atrial fibrillation.

3) Timing of predictor measurement of candidate predictors (e.g. at patients presentation, at diagnosis, at treatment initiation)

- DECAF indices were assessed on admission (see table 1).

4) Were predictors assessed blinded for outcome, and for each other (if relevant)?

- Predictors were assessed blinded from the outcome. The external validation cohort was identified prospectively, so variables were collected prior to the outcome. The internal validation study was performed retrospectively. Three DECAF variables, eosinopenia, acidaemia and atrial fibrillation, are objective. Potentially, there may be a degree of inter-observer variation in the reporting of chest radiograph consolidation and the eMRCD score, however the research team relied on the observations of the attending clinicians. For patients identified retrospectively, the researcher

obtaining the information from the notes was blinded to the outcome. Collection of individual predictors was not blinded from other DECAF indices, although the consequent risk of bias is low.

5) Handling of predictors in the modelling (e.g. continuous, linear, non-linear transformations or categorised)

- The DECAF variables were applied as per the derivation study. eMRCD score is categorised, eosinophil score and pH are dichotomised, and AF and chest x-ray consolidation are binary. Dichotomising variables can cause a loss of discrimination, depending on their relationship with the outcome. This was not an issue as the continuous variables related to DECAF show a non-linear relationship to mortality, and the pre-define thresholds were optimal. Discrimination of DECAF was very good, and similar to that of the derivation study, in both validation cohorts.

Sample size

1) Number of participants and number of outcomes/events

- The internal and external cohorts were individually adequately powered. Internal cohort: 880 participants, 78 events; external cohort 845 participants, 54 events.

2) Number of outcomes/events in relation to the number of candidate predictors (events per variable)

- This approach to sample size calculation is only relevant to the derivation study.

Missing data

1) Number of participants with any missing value (include predictors and outcomes)
2) Number of participants with missing data for each predictor

- 1+2) Number of missing values and number of participants with missing data provided.

3) Handling of missing data (e.g. complete-case analysis, imputation, or other methods)

- 3) Low rates of missing data. Multiple imputation used; complete-case analysis also performed.

Model development

1) Modeling method (e.g. logistic. survival, neural networks, or machine learning techniques)
2) Modeling assumptions satisfied
3) Methods for selection of predictors for inclusion in multivariable modelling (e.g. all candidate predictors, pre-selection based on unadjusted association with the outcome)
4) Method for selection of predictors during multivariable modeling (e.g. full model approach, backward or forward selection) and criteria used (e.g. p-value, Akaike Information Criterion)

- 1-4) The DECAF model was developed in the DECAF derivation study; these aspects do not apply to the validation study.

5) Shrinkage of predictor weights or regression coefficients (e.g. no shrinkage, uniform Shrinkage, penalized estimation)

- 5) Shrinkage refers to adjusting coefficients to protect against overfitting and loss of discrimination in validation studies. In developing the DECAF score, the prognostic indices were weighted based on their coefficients. The same weighting was used for both validation cohorts, and discrimination remained good in both validation cohorts.

Model performance

1) Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and discrimination (C-statistic, D-statistic, log rank) measures with confidence intervals

- Hosmer-Lemeshow test is provided, and observed risk from derivation and validation cohorts described and compared. The validation study showed good calibration. Although the absolute risk differed between the derivation and validation study for high risk patients, this reflects differences in overall mortality rates and a large and stepwise increase in mortality is seen between different risk groups.

2) Classification measures (e.g. sensitivity, specificity, predictive values, net reclassification improvement) and whether a priori cut points were used

- 2) Sensitivity and specificity are provided. Reclassification measures, such as net reclassification improvement, look at the value in adding a single predictor to a prediction model. Due to the very strong performance of the DECAF score, no reclassification measures were performed or required.

Model evaluation

1) Methods used for testing model performance: development dataset only (random split of data, resampling methods, e.g. bootstrap or cross-validation, none) separate external validation (e.g. temporal, geographical, different setting, different investigators)

- Internal validation: involved the same hospitals as the derivation study, but at a different time period (a form of temporal validation) and additional investigators.
- External validation was performed in hospitals that are geographically distinct. Hospitals were chosen to ensure variation in population characteristics (rurality and socioeconomic factors) and structures of care to maximise generalisability. The research staff within external sites were not involved in the derivation study.

2) In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)

- Not applicable

Results

1) Final and other multivariable models (e.g. basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard or confidence Intervals)
2) Any alternative presentation of the prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance
3) Comparison of the distribution of predictors (including missing data) for development and validation datasets

- 1+2+3) Predictor weights and regression coefficients are given for the DECAF score. All models have AUROC calculated with confidence intervals. Subgroup with pneumonia presented. As with the original DECAF study, missing data rates for both validation cohorts was low.

Interpretation and discussion

1) Interpretation of presented models (confirmatory, if model useful for practice versus exploratory, is more research needed)

- The performance of DECAF is excellent in two separate, and individually adequately powered, validation cohorts. This confirms that DECAF can risk stratify patients effectively. The value of using DECAF to inform clinical practice, such as to identify patients for Hospital at Home, requires further research.

2) Comparison with other studies, discussion of generalisability strengths and limitations.

- DECAF is compared to other prognostic scores, with discussion of the strengths and limitations.