Early View

Original article

The NIV Outcomes (NIVO) Score: prediction of in-hospital mortality in exacerbations of COPD requiring assisted ventilation

Tom Hartley, Nicholas D. Lane, John Steer, Mark W. Elliott, Milind P. Sovani, Helen Jane Curtis, Elizabeth R. Fuller, Patrick B. Murphy, Dinesh Shrikrishna, Keir E. Lewis, Neil R. Ward, Chris D. Turnbull, Nicholas Hart, Stephen C. Bourke

This manuscript has recently been accepted for publication in the *European Respiratory Journal*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJ online.

The NIV Outcomes (NIVO) Score: prediction of in-hospital mortality in exacerbations of COPD requiring assisted ventilation.

Author	Affiliations	Address
Tom Hartley	1. Royal United Hospital Bath NHS Foundation Trust* 2. Translational and Clinical Research Institute, Newcastle University (* TH was employed by Northumbria Healthcare NHS Foundation trust throughout the NIVO study)	Royal United Hospital, Combe Park, Bath, BA1 3NG
Nicholas D Lane	1. Northumbria Healthcare NHS Foundation Trust 2. Translational and Clinical Research Institute, Newcastle University	Respiratory Department, North Tyneside General Hospital, North Shields, Tyne and Wear, NE29 8NH
John Steer	1. Northumbria Healthcare NHS Foundation Trust 2. Translational and Clinical Research Institute, Newcastle University	Respiratory Department, North Tyneside General Hospital, North Shields, Tyne and Wear, NE29 8NH
Mark W Elliott	1. The Leeds Teaching Hospital NHS Trust.	St James's University Hospital Leeds, LS9 7TF
Milind P Sovani	1. Nottingham University Hospitals NHS Trust	Nottingham University Hospitals NHS Trust D floor South Block Queens Medical Centre Nottingham, NG7 2UH
Helen Jane Curtis	1. Gateshead Health NHS Foundation Trust	Queen Elizabeth Hospital, Gateshead, Tyne and Wear, NE9 6SX
Elizabeth R Fuller	1. South Tyneside and Sunderland NHS Foundation Trust 2. Population Health Sciences Institute, Newcastle University	South Tyneside District General Hospital, South Shields, UK
Patrick B Murphy	1. Guy's & St Thomas' NHS Foundation Trust 2. Centre for Human and Applied Physiological Sciences, King's College London, UK	Lane Fox Respiratory Service, Guy's & St Thomas' NHS Foundation Trust, London, UK
Dinesh Shrikrishna	1. Musgrove Park Hospital, Somerset NHS Foundation Trust	Musgrove Park Hospital, Parkfield Drive, Taunton, TA1 5DA
Keir E Lewis	1. School of Medicine, University of Swansea 2. Hywel Dda University Health Board	Respiratory Unit, Prince Philip Hospital, Llanelli, Wales. S14 8QF

Neil R Ward	1. University Hospitals Plymouth NHS Trust	Chest Clinic, Derriford Hospital, Plymouth, UK. PL6 8DH.
Chris D Turnbull	1. Nuffield Department of Medicine, University of Oxford 2. Oxford University Hospitals NHS Foundation Trust	Oxford Centre for Respiratory Medicine, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LE
Nicholas Hart	1. Guy's & St Thomas' NHS Foundation Trust 2. Centre for Human and Applied Physiological Sciences, King's College London, UK	Lane Fox Respiratory Service, Guy's & St Thomas' NHS Foundation Trust, London, UK
Stephen C Bourke	1. Northumbria Healthcare NHS Foundation Trust 2. Translational and Clinical Research Institute, Newcastle University	Respiratory Department, North Tyneside General Hospital, North Shields, Tyne and Wear, NE29 8NH

The NIV Outcomes (NIVO) Score: prediction of in-hospital mortality in exacerbations of COPD requiring assisted ventilation.

Abstract

Introduction: Acute exacerbations of COPD (AECOPD) complicated by acute (acidaemic) hypercapnic respiratory failure (AHRF) requiring ventilation are common. When applied appropriately, ventilation substantially reduces mortality. Despite this, there is evidence of poor practice and prognostic pessimism. A clinical prediction tool could improve decision making regarding ventilation, but none is routinely used.

Methods: Consecutive patients admitted with AECOPD and AHRF treated with assisted ventilation (principally non-invasive ventilation) were identified in two hospitals serving differing populations. Known and potential prognostic indices were identified a priori. A prediction tool for in-hospital death was derived using multivariable regression analysis. Prospective, external validation was performed in a temporally separate, geographically diverse 10-centre study. The trial methodology adhered to TRIPOD recommendations.

Results: Derivation cohort, n=489, in-hospital mortality 25.4%; validation cohort, n=733, in-hospital mortality 20.1%. Using 6 simple categorised variables; extended Medical Research Council Dyspnoea score (eMRCD)1-4/5a/5b, time from admission to acidaemia >12 hours, pH <7.25, presence of atrial fibrillation, Glasgow coma scale ≤14 and chest radiograph consolidation a simple scoring system with strong prediction of in-hospital mortality is achieved. The resultant NIVO score had area under the receiver operated curve of 0.79 and offers good calibration and discrimination across stratified risk groups in its validation cohort.

Discussion: The NIVO score outperformed pre-specified comparator scores. It is validated in a generalisable cohort and works despite the heterogeneity inherent to both this patient group and this intervention. Potential applications include informing discussions with patients and their families, aiding treatment escalation decisions, challenging pessimism, and comparing risk-adjusted outcomes across centres.

Introduction

Acute Exacerbations of COPD (AECOPD) account for over 141,000 admissions per year in the UK [1], of which a quarter are complicated by (acidaemic) acute hypercapnic respiratory failure (AHRF) during hospital admission [2, 3]. When acidaemia occurs, guidance unambiguously supports the use of non-invasive ventilation (NIV) [4, 5]. A Cochrane review states; NIV reduces need for intubation by 65% (NNT 5), reduces mortality by 46% (NNT 12) and further trials to prove efficacy are unwarranted [6]. However, NIV is often underused and/or poorly delivered. UK national audit data suggests more than half of patients with AECOPD and AHRF do not receive NIV, yet only around 20% will correct with medical therapy alone [2, 3, 7]. Most clinicians' estimates of outcome have been shown to be pessimistic which may contribute to underuse [8]. Compounding the problem, acute NIV services are arguably under-developed compared to other services delivering emergency medical intervention such as for stroke or myocardial infarction. These concerns, and substantial variation by institution, were noted in the National Confidential Enquiry into Patient Outcome and Death (NCEPOD) review of national practice in the UK [9]. Such recent, published, national data is less robust outside of the UK but there is evidence of substantial variation in practice in both Europe and North America [10-13].

The decision to instigate ventilation is complex, should involve clinicians with suitable expertise, and requires detailed knowledge of several factors including the timing of initiation, the magnitude of treatment effect and the overall likelihood of successful treatment. These factors should be communicated with a patient or family to reach a shared decision. In a recent review of UK practice, more than two thirds of decisions to initiate NIV were made by a non-specialist below consultant grade and the first senior review was by a specialist (respiratory or intensive care medicine) in only 31.7% of cases [9].

A countermeasure to poor prognostic accuracy is objective stratification of patient outcomes to allow clinicians and patients better understanding of the likely success of the intervention. To our knowledge, no current predictive model is in widespread use in this setting. The most focussed previous attempt was reported by Confalonieri et al but requires computation and derives much of its predictive power from events after the introduction of ventilation and therefore is not useful to guide the initial decision [14]. In keeping with many predictive models, it has relatively limited external validation. Other predictive tools such as the APACHE II [15], COPD and Asthma Physiology Score (CAPS) [16], CURB 65 [17],

HACOR [18], or DECAF [19, 20] score may potentially be used to augment decision making. APACHE II and CAPS are complex, were derived in selected populations and rely on information not always available at the bedside. CURB 65 is simple to apply, but was not designed for use in this population. HACOR is recently reported and therefore not pre-specified in our protocols, but we have assessed its performance relative to other tools. It predicts NIV failure after introduction of NIV rather than mortality. The DECAF score (developed by some members of the NIVO research group) predicts in-hospital mortality in all exacerbations of COPD whether or not acidaemia is present; the AUROC is 0.82-0.86 overall, but performance is less strong in patients requiring ventilation.

The absence yet potential value of, a good predictive model has been noted frequently including by the Cochrane consortium *"additional research would enhance our ability to more accurately select the right patients and the right levels of ventilation"* [6]. A well-constructed predictive model has potential to challenge pessimism, standardise practice, inform discussions with patients about their care, guide level of care and facilitate audit.

Methods

Programme Overview

The Non-Invasive Ventilation Outcomes (NIVO) programme aimed to derive and validate a tool to predict outcome in AECOPD complicated by AHRF. The guiding principles of this work are that it adhere to TRIPOD methodology, [21] the population(s) be generalisable and that the resultant tool be simple, have face validity and use only data immediately available to the attending clinician.

NIVO has two distinct parts; firstly, a retrospective study in two hospitals with diverse catchment areas and structures of care within a single trust (Northumbria Healthcare NHS Foundation Trust) to derive and develop a predictive model. Secondly, a 10-centre, prospective validation study to test the findings in a separate population. Sites here are reported as A-J to maintain anonymity (although we report the lead site is site A). Prior to commencement, a patient group discussed, amended and endorsed proposed methods. Both derivation and validation protocols were published prior to recruitment, ISRCTN 16977236 and 22921168 respectively. In this observational study usual care was unaffected and hence individual patient consent was not required; ethical approval was obtained from NHS regional ethics boards. Particular attention was paid to ensuring that the study population was

generalisable and that consecutive patients were identified. External validation sites were selected to ensure few patients are denied ventilation on assumption of futility and to maximise geographical diversity in hospitals of varying size. Further enhancing validation, the lead site underwent wholesale reorganisation to the model of acute care delivery between the derivation and validation cohorts.

Inclusion criteria were: AECOPD as primary diagnosis, preadmission spirometry evidence of airflow obstruction (forced expiratory volume in one second (FEV_1)/ vital capacity (VC) <0.7), AHRF (defined by time matched guideline) treated with assisted ventilation (NIV or invasive mechanical ventilation (IMV)), smoking history of ≥10 pack years and age ≥35 years. Exclusion criteria were previous inclusion in study or illness other than COPD likely to limit life to less than one year (principally metastatic cancer).

Data Collection.

The derivation cohort was identified by cross referencing pre-existent databases, including NIV rolling audit, with coding searches to ensure capture of consecutive unique patients. Demographics, population and clinical descriptors, haematological/biochemical/arterial blood gas (ABG) data, components of comparator models and indices associated with mortality from literature review were collected. Each case record (paper notes and electronic records) were reviewed, and a bespoke database compiled. Presence of chest radiograph consolidation was determined using the following hierarchy: attending senior clinician interpretation, radiologist report, researcher interpretation.

To standardise the recording of dynamic variables (for example, physiological data), the value with greatest deviation from normal in the 24 hours prior to the decision to instigate ventilation was collected.

In the validation cohort, data collection was limited to descriptors, components of relevant published prognostic tools (Confalonieri risk chart, APACHE II, COPD and Asthma Physiology Score (CAPS), DECAF score, CURB-65 and any index from the derivation study with a univariate association with mortality. Although not prespecified the recently developed HACOR score is also compared. Patients were prospectively identified by daily screening of locations delivering ventilation; ventilation service records and coding searches were used to ensure no potentially eligible patients were missed. Lead site recruitment was limited to 200 patients.

Statistical Methods

To estimate the sensitivity of the tool (assuming expected sensitivity of 70% and standard error of 5%) 85 deaths should be studied in each cohort. With an estimated in-hospital mortality rate of 20% at least 425 patients are required in both the derivation and validation cohorts.

For variables with <20% missing, data was assumed to be missing at random and imputed using the expectation-maximisation (EM) algorithm. Data was characterised into mean + standard deviation (SD) if parametric, median + interquartile range (IQR) if non-parametric and as a proportion if categorical. Univariate analysis was performed using Student's T test, Mann-Whitney U test and Chi-Squared test respectively. Multicollinearity between potential predictor variables were handled according to recommendations by Field [22].

To build the prognostic tool the following steps were followed:

A. Potential predictor variables were determined: univariate association with mortality (p <0.1).

B. Collinearity was assessed; where present the variable with strongest plausibility or most significant association was used.

C. Variables were considered for further assessment provided there was a plausible or established association with mortality, missing data was <10% and, if categorical, were not highly asymmetrically split (>90:10).

D. Multivariable analysis; logistical regression with a backward, stepwise entry method was used.

E. The remaining continuous independent predictors in the resultant 'full model' were simplified. Dichotomy used the following hierarchical approach: area under the receiver operated curve (AUROC) analysis, results from previous research, a clinically meaningful value or a median split. The extended medical research council dyspnoea scale (eMRCD) has 3 categoriesin line with previous research [19, 20].

F. Categorised variables were re-entered into regression analysis to ensure they retained prognostic value. The strongest remaining variables according to their regression co-efficient were included and relative weighting was ascribed [23].

G. Calibration and model fit were assessed using calibration plot, Hosmer-Lemeshow goodness of fit test and R^2. Studentised residuals and Cook's distance allowed evaluation of outliers.

H. Discrimination of final model and comparison to other models was by AUROC analysis and statistical difference assessed using the method of Delong and colleagues. [24] Performance of the tool using categorised (simple model) and continuous (full regression equation) indices was compared.

I. Predictor variables and weightings from derivation model were examined in the validation cohort to determine whether further simplification was feasible.

For the validation cohort, data was handled in the same manner. Sites did not know which of the many indices they collected were the predictors of outcome to prevent knowledge influencing usual care. Anonymised data was submitted via a digital platform as close to real time as possible with continuous remote data monitoring and a final data monitoring visit.

Throughout univariate analysis and during modelling we used in-hospital mortality as the dependent variable, as per protocol. 90-day outcomes are also reported.

To aid clinical decision making, we also examined whether a small number of strong clinical predictors could be used to identify a particularly high risk cohort in whom ventilation may not be in the patients best interests, termed 'rule of thumb.

Results

Derivation cohort.

489 unique, consecutive patients meeting selection criteria were admitted between 30/11/08 and 19/5/13 to Northumbria Healthcare NHS Foundation Trust. 124 of 489 (25.4%) died in hospital. Missing data is shown in the online supplement, Table 1 shows population descriptors. Admissions were split between the two Trust sites (52% vs 48%). Current smokers represented 48.7%, and total smoking burden was high; mean (SD) 49.5 (26.0)

cigarette pack years. 70.1% were admitted from their own home without a formal care package. Median (IQR) eMRCD score was 5a (4-5b) suggesting a substantial number of housebound patients who may be in receipt of informal care. Following admission, most patients (94.5%) received only NIV, the remainder received IMV +/- NIV. Chest X-ray consolidation was present in 47.2%.

Time from admission to index episode of acidaemia was 146 (56-852) minutes with 73.8% occurring in the first 12 hours. At ventilation initiation, median IPAP/EPAP was 16/4 cmH$_2$0 rising to 18/4 at 1 hour and 20/4 maximum. Amongst survivors, median (IQR) length of stay was 10 (7-17) days; time to inpatient death was 7 (2-14) days.

Tool Building

Following methodological steps A to C, 21 variables were entered into the regression equation (see online supplement). Univariate significance was unchanged with or without imputed data. Some statistical associations with mortality were rejected: Albumin (missing data rates), Mean arterial pressure and BMI (not routinely available at the bedside), admission from institutional care (multiple confounders), ineffective cough (subjective assessment required) and current smoker (protective and associated with multiple confounders). Of note, no lung function measurement was associated with in-hospital mortality. Detailed comorbidity information was collected but only left ventricular systolic dysfunction and atrial fibrillation showed univariate association with in-hospital mortality.

Following step D (regression using continuous variables), 11 variables remained. Step E generated categorical variables and multivariable regression was repeated (step F): table 2. Of note, while pH and base excess did not meet definitions of collinearity, they measure similar concepts and interact when model building. Due to the strong face validity argument in favour of pH and to avoid overfitting to a derivation dataset, we chose to include pH rather than the marginally stronger base excess (negative BE associated with mortality). Similarly, in choosing our 8 variables to form the derivation model, we chose pH over respiratory rate as respiratory rate is highly dynamic and related to frequency of observation.

A simple tool to predict in-hospital mortality, scoring 1 point for each variable and 2 for an eMRCD score of 5b, yielded an area under the receiver operated curve (AUROC) in the derivation cohort of 0.85 (0.82-0.89).

Table 2 <u>Derivation Cohort</u>: Multi-variable, logistical regression results using in-hospital mortality as the dependent variable.

	B	Significance	(Odds Ratio 95% CI)
Chest radiograph consolidation	1.019	<0.0001	2.77 (1.61-4.76)
Glasgow Coma Scale (GCS) ≤14	0.803	0.004	2.23 (1.29-3.87)
Atrial fibrillation present	1.298	<0.0001	3.66 (1.93-6.96)
pH <7.25	0.571	0.042	1.77 (1.02-3.07)
Time to acidaemia >12 hours	1.484	<0.0001	4.41 (2.49-7.80)
eMRCD 5a	1.159	0.001	3.19 (1.67-6.07)
eMRCD 5b	1.981	<0.0001	7.25 (3.50-15.03)
Long term Oxygen Therapy (LTOT)	0.764	0.012	2.15 (1.19-3.88)
Respiratory rate ≥30	0.675	0.012	1.97 (1.16-3.33)
Eosinophil count <0.05 x10^9/L	1.538	<0.0001	4.66 (2.51-8.64)
Intercept -4.619, R^2 0.465, Hosmer and Lemeshow 0.262. Cook's distances and Studentised residuals all within acceptable limits.			

Table 1 Key population descriptors.

	Derivation Cohort	Validation Cohort	Individual validation sites									
			A	B	C	D	E	F	G	H	I	J
N	489	733	200	116	77	69	67	60	49	44	37	14
Female (%)	62.6	58.3	56.5	62.9	62.3	63.8	53.7	50	46.9	63.6	56.8	78.6
Age (years)	72.8 (10.0)	70.5 (9.3)	71.9 (9.2)	68.9 (8.5)	70.2 (10.1)	72.8 (10.1)	70.5 (9.6)	67.7 (9.1)	71.8 (8.6)	70.1 (9.6)	68.9 (8.3)	68.3 (10.5)
BMI	24.6 (7.3)	25.5 (8.0)	25.9 (8.3)	23.2 (6.7)	26.9 (8.4)	24.8 (3.5)	28.1 (9.6)	25.5 (7.8)	25.7 (7.1)	26.7 (7.7)	24.5 (7.3)	21.1 (6.2)
eMRCD score	5a (4-5a)	5a (4-5a)	5a (4-5a)	5a (4-5a)	5a (4-5a)	4 (4-5a)	5a (4-5b)	4 (4-5a)	4 (4-5a)	4 (4-5a)	5a (4-5b)	3 (2-4)
FEV$_1$%	38.0 (16.4)	37.2 (15.4)	40.5 (16.6)	34.5 (14.6)	38.6 (13.3)	38.1 (16.2)	38.1 (15.5)	30.6 (13.8)	35.8 (13.4)	36.5 (13.4)	36.9 (16.7)	34.0 (13.8)
LTOT (%)	29.2	28.6	25	30.2	26	33.3	35.8	38.3	24.5	15.9	35.1	21.4
Prev NIV (%)	21.9	35.9	40	37.1	28.6	29	34.4	58.3	28.6	20.5	37.8	21.4
HMV (%)	2.0	8.7	5.5	10.3	9.1	4.3	4.5	3.0	6.1	2.3	13.5	7.1
pH	7.26 (7.20-7.30)	7.26 (7.21-7.30)	7.27 (7.22-7.30)	7.26 (7.21-7.29)	7.26 (7.19-7.29)	7.27 (7.21-7.29)	7.30 (7.23-7.32)	7.26 (7.22-7.29)	7.23 (7.17-7.27)	7.27 (7.20-7.30)	7.25 (7.17-7.27)	7.21 (7.16-7.29)
CO$_2$ (kPa)	9.9 (8.5-11.7)	10.2 (2.7)	10.1 (2.7)	9.9 (2.6)	10.3 (2.2)	10.1 (3.5)	10.1 (2.6)	10.0 (2.4)	10.6 (2.2)	10.1 (2.5)	11.7 (3.3)	11.5 (3.2)
Max IPAP (cmH$_2$O)	20 (18-20)	20 (18-24)	24 (22-26)	20 (15-22)	17 (14-20)	20 (16-22)	20 (16-22)	21 (17-27)	20 (16-20)	22 (17-27)	20 (20-25)	20 (14-20)
IMV (%)	5.5	2.9	2	1.7	1.3	5.8	4.5	8.3	0	2.3	2.7	0
APACHE II score	20 (16-23)	19 (16-22)	19.5 (15-23)	18 (16-22)	20 (18-23)	19 (16-22)	18 (14-20)	18.5 (14-22)	18 (16-23)	18 (14-21)	21 (18-24)	16 (14-23)

Body Mass Index (BMI). Percentage of predicted forced expiratory volume in 1 second (FEV$_1$%). Long term oxygen therapy (LTOT). Previous non-invasive ventilation (Prev NIV). Home mechanical ventilation (HMV). pH at ventilation (pH). Arterial carbon dioxide at ventilation (CO$_2$ kPa). Maximum achieved inspiratory positive airway pressure (Max IPAP). Received invasive mechanical ventilation (IMV).

<u>Validation cohort.</u>

Recruitment was open between 14/10/16 and 28/02/18, although not all sites recruited for the whole period. 733 unique, consecutive patients admitted to centres A-J were prospectively recruited. In-hospital mortality was 20.1%, and 90-day mortality was 32.2%. The missing data rate was low; amongst the 8 variables from the derivation project missing data was 0% except GCS (2.3%) and eosinophil count (2.9%) where data was inputted using the expectation-maximisation algorithm (see online supplement for full details).

The highest level of care for each patient was recorded with in-hospital mortality rate shown in brackets: 46.7% managed on respiratory support unit (17.3% mortality), 32.2% (22.5% mortality) medical ward, 13.5% (20.2% mortality) high dependency unit, 7.6.% (26.8% mortality) intensive care unit. Place of care was not significantly associated with in-hospital mortality. Median length of stay was 8 (IQR 6-14) days. 98% of participants were Caucasian (95.5% over 65s in England and Wales are Caucasian) [25]. Significant antecedent tobacco burden was seen, mean 44.8 pack years (SD 23.7). 68.9% had been admitted to hospital for any reason in the preceding year.

Table 3 <u>Validation Cohort</u>: Multi-variable, logistical regression results using in-hospital mortality as the dependent variable.

	B	Significance	Odds Ratio (95% CI)
Chest radiograph consolidation	0.358	0.089	1.43 (0.95-2.16)
Glasgow Coma Scale (GCS) ≤14	0.658	0.002	1.93 (1.26-2.95)
Atrial Fibrillation	0.842	<0.0001	2.32 (1.45-3.71)
pH <7.25	0.961	<0.0001	2.61 (1.69-4.04)
Time to Acidaemia >12 hours	1.289	<0.0001	3.63 (2.33-5.64)
eMRCD 5a	1.425	<0.0001	4.16 (2.46-7.02)
eMRCD 5b	1.960	<0.0001	7.10 (4.05-12.46)

Intercept -2.832, R^2 0.285, Hosmer and Lemeshow 0.130. Leverage values, Cook's distances and Studentised residuals all within acceptable limits.

The 8 variables from the derivation study were examined (using logistic regression analysis) to see if further simplification was possible (see online supplement). Of importance, identical

variables using the pre-determined cut points from the derivation study were assessed and no further exploration of validation data for novel associations with mortality was performed. This showed that both the eosinophil count and long-term oxygen prescription could be removed (non-significant in multivariable regression) from the model without detrimental effect. Table 3 shows final regression data. Therefore, in keeping with pre-determined aim to create the simplest model, a 6 variable model was finalised. Following the removal of variables, re-weighting was applied; of note reweighting was unchanged whether determined by the derivation or validation cohort. The model, termed the NIVO score, with variable weightings is shown in Figure 1. Figure 2 shows comparative performance.

Figure 1 The NIVO score.

Figure 2

Table 4 shows the in-hospital and 90-day mortality by NIVO score and simple risk categories have been created. The NIVO score significantly outperforms prespecified comparator scores when predicting in-hospital mortality (p< 0.001, NIVO vs each other tool, using method described by Delong). [24] Moreover, if continuous variables were not categorial transformed the model accuracy is not substantially improved: AUROC 0.80 (0.76-0.84).

Table 4 In-Hospital and 90-day mortality by NIVO score increment and risk category.

NIVO Score	N	In-Hospital Mortality	90-day mortality
0	67	0%	10.4%
1	79	8.9%	20.3%
2	133	5.3%	15.8%
3	152	15.1%	26.3%
4	116	19.0%	40.5%
5	97	35.1%	46.4%
6	54	53.7%	59.3%
7	26	65.4%	76.9%
8	8	87.5%	87.5%
9	1	100%	100%
Total	733	20.1%	32.2%
Risk Category			
Low (0-2)	279	5.0%	15.8%
Medium (3-4)	268	16.8%	32.5%
High (5-6)	151	41.2%	50.1%
Very High (7-9)	35	71.4%	80.0%

Table 5 shows very simple 'rules of thumb' employing only 2 variables to identify particularly high-risk patients.

Table 5 Rules of thumb.

Rule of thumb	DERIVATION In-hospital mortality	VALIDATION In-hospital mortality	VALIDATION 90 day mortality
MRCD 5 (eMRCD 5a/5b) + acidaemia >12 hours after admission	46/77 (59.7%)	50/101 (49.5%)	57/101 (56.4%)
MRCD 5 (eMRCD 5a/5b) + acidaemia >48 hours after admission	27/40 (67.5%)	30/45 (66.7%)	36/45 (80.0%)

Discussion

This project represents the complete derivation and prospective, multi-centre validation of a predictive model, the NIVO score; with which patients can be accurately stratified according to the risk of in-hospital mortality.

Nearly 40% of patients (270/733) fall into the low risk group with a low (5.0%) in-hospital mortality. This group also have a much lower 90-day mortality than other patients. At the other end of the spectrum, the NIVO score allows for identification of a smaller cohort of patients with very high in-hospital and 90-day mortality.

It is important to note what is not included in the NIVO score. Certain predictors with little association with outcome may be being used to inappropriately support pessimistic practice; neither FEV_1 nor routinely measured blood tests are important. Others, such as age and LTOT prescription are of limited use. Magnitude of pH derangement and presence of consolidation are important but less than expected due to prediction being spread across multiple variables.

Strengths: Validation was prospective and temporally separated from derivation; it encompasses different structures of care in both large and small hospitals with a geographical distribution of sites throughout England and Wales. Considerable effort was given to ensuring included patients are consecutive to avoid survival bias. The datasets were generated specifically for this project and guided by extensive literature review and collated expert opinion. This approach allows study of a breadth of indices and avoids the limitations associated with mining an existing dataset when many potentially interesting candidate predictors are not available. Objective verification of smoking history and airflow obstruction combined with researcher determination that the reason for admission was an exacerbation of COPD confers substantial advantage over coding-based strategies by eliminating AECOPD mimics. Similarly, by not imposing many other selection criteria, a population that is representative of real-world patients was included and is hence readily generalisable. Representing real world practice, many patients in the study had previously been treated with NIV at the time of their index admission. The inclusion of NIV naïve, those with previous acute exacerbations requiring NIV and a smaller number in receipt of domiciliary ventilation maximises the utility of the NIVO score.

Weaknesses: Some may contend that concurrent pneumonia invalidates the diagnosis of an exacerbation of COPD, however this is not the study group's position [26]. Patients with X-ray consolidation were included in many of the trials originally investigating NIV in AECOPD. It is our contention that considering radiographic consolidation as an adverse marker is the more logical approach. Supporting this, in a small RCT of NIV in pneumonia a survival benefit was only seen in the subgroup with underlying COPD [27].

We acknowledge several potential weaknesses, albeit conscious ones. In not controlling the intervention, there may be a cohort of patients that met guideline criteria for ventilation but did not receive it. We deliberately included sites with well-established ventilation services and scrutinised audit data prior to acceptance to mitigate for this. If excessive patient selection had taken place then mortality would likely be lower, and the population described less disabled (median patient is housebound, eMRCD 5a). Secondly, the intervention has been heterogeneously applied. Protocolising inclusion and intervention would be less representative of usual clinical care, consequently in this specific observational study the heterogeneity and generalisability of the population may have been lost. The superior performance of the NIVO score over comparators within the uncontrolled population supports its use in routine clinical practice.

Some influential groups have reported opinion on best practice for prognostic modelling [28]. For the most part they advocate a parsimonious approach particularly in handling of continuous variables, i.e. to maintain variables as continuous or use deciles. This leads to complex scores such as the APACHE II score. We have adopted an approach that values simplicity in the assumption that overly complex tools are rarely adopted into mainstream use and therefore unlikely to lead to patient benefit. Nevertheless, we publish the full regression results, which in an era of increasing computation may become useful. It is important to emphasise that when the continuous variables in the NIVO score were categorically transformed, performance was relatively unaffected, adding credence to our approach. The likely explanation is that incremental risk across the range of various indices is not linear, but rather markedly skewed meaning little predictive power is lost by categorising variables.

Comparison to previous research: The NIVO score significantly outperforms all comparison scores, has face validity and produces clinically meaningful risk stratification. By using only 6 simple and readily available variables not only can the likelihood of in-hospital mortality be predicted but there is also strong correlation with 90-day mortality.

Following literature review, the eMRCD score and the time to development of acidaemia were of particular interest as predictors of outcome in this setting (the importance of timing has since the original literature search again been highlighted) [20, 29, 30]. These variables had not been included in a model to predict outcome in this setting before and were the strongest predictors of outcome. They are likely to account for outperformance of more complex scores (APACHE II and CAPS) and furthermore, why the next best comparator (DECAF which employs the eMRCD) offers reasonable performance [20]. The recently reported HACOR score includes PaO2/FiO2 ratio. Whether tested in the whole population using estimated values from uncontrolled oxygen or limited to those with known FiO2, NIVO is markedly better. The rules of thumb (table 5) further explore simple ways to prognosticate using these indices alone. Combined, they can identify small numbers of higher-risk patients. This has the benefit of being easy to remember and may assist decision making regarding treatment escalation.

Potential uses and future research: Clinical tools do not replace individualised decision making but add valuable supportive data. Clinicians can be inaccurate and pessimistic when predicting outcome in this setting. If adopted, the NIVO score could better objectify expected outcome, and challenge pessimism, improving timely provision of NIV when indicated. Stratified risk has potential to improve standardisation around decisions such as entry to higher level of care beds or be employed in national audit programmes to facilitate comparisons between units.

Shared decision making is an important aspect of modern healthcare, crucial to this is communication of fact. More objective assessment can contribute to truly shared decisions. Linked to this, NIV can be an intrusive treatment, poorly tolerated by some. Identification of those at greatest risk of death could help inform the decision between clinician, patient and family to instigate palliative care in lieu of active treatment options.

Conclusion

The NIVO score allows for accurate risk stratification of patients admitted to hospital with AECOPD complicated by acidaemia and AHRF who required assisted ventilation. It does so using simple, readily available information, and is generalisable to real world conditions. In this common condition, poor practice is widespread despite an excellent treatment; we

foresee the NIVO tool's greatest strength is in challenging pessimism and increasing timely access to lifesaving treatment.

References

1.   British Lung Foundation. Lung disease in the UK – big picture statistics [Available from: https://statistics.blf.org.uk/lung-disease-uk-big-picture.

2.   Royal college of Physicians. Report of The National Chronic Obstructive Pulmonary Disease Audit 2008: clinical audit of COPD exacerbations admitted to acute NHS units across the UK. Royal College of Physicians, British Thoracic Society; 2008.

3.   Royal college of Physicians. COPD: Who Cares Matters. National Chronic Obstructive Pulmonary Disease (COPD) Audit Programme: Clinical audit of COPD exacerbations admitted to acute units in England and Wales 2014. Royal College of Physicians, British Thoracic Society; 2015.

4.   Davidson AC, Banham S, Elliott M, Kennedy D, Gelder C, Glossop A, Church AC, Creagh-Brown B, Dodd JW, Felton T, Foëx B, Mansfield L, McDonnell L, Parker R, Patterson CM, Sovani M, Thomas L, BTS Standards of Care Committee Member BTSICSAHRFGDG, On behalf of the British Thoracic Society Standards of Care Committee. BTS/ICS guideline for the ventilatory management of acute hypercapnic respiratory failure in adults. Thorax. 2016;71(Suppl 2):ii1-ii35.

5.   Rochwerg B, Brochard L, Elliott MW, Hess D, Hill NS, Nava S, Navalesi P, Antonelli M, Brozek J, Conti G, Ferrer M, Guntupalli K, Jaber S, Keenan S, Mancebo J, Mehta S, Raoof S. Official ERS/ATS clinical practice guidelines: noninvasive ventilation for acute respiratory failure. European Respiratory Journal. 2017;50(2):1602426.

6.   Osadnik CR, Tee VS, Carson-Chahhoud KV, Picot J, Wedzicha JA, Smith BJ. Non-invasive ventilation for the management of acute hypercapnic respiratory failure due to exacerbation of chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. 2017(7).

7.   Plant PK, Owen JL, Elliott MW. Early use of non-invasive ventilation for acute exacerbations of chronic obstructive pulmonary disease on general respiratory wards: a multicentre randomised controlled trial. The Lancet. 2000;355(9219):1931-5.

8.   Wildman MJ, Sanderson C, Groves J, Reeves BC, Ayres J, Harrison D, Young D, Rowan K. Implications of prognostic pessimism in patients with chronic obstructive pulmonary disease (COPD) or asthma admitted to intensive care in the UK within the COPD and asthma outcome study (CAOS): multicentre observational cohort study. BMJ. 2007;335(7630):1132.

9.	The national confidential enquiry into patient outcome and death (NCEPOD). Inspiring change. A review of the quality of care provided to patients receiving acute non-invasive ventilation.; 2017.

10.	Crimi C, Noto A, Princi P, Esquinas A, Nava S. A European survey of noninvasive ventilation practices. European Respiratory Journal. 2010;36(2):362-9.

11.	Bierer GB, Soo Hoo GW. Noninvasive Ventilation for Acute Respiratory Failure: A National Survey of Veterans Affairs Hospitals. Respiratory Care. 2009;54(10):1313-20.

12.	Chandra D, Stamm JA, Taylor B, Ramos RM, Satterwhite L, Krishnan JA, Mannino D, Sciurba FC, Holguín F. Outcomes of Noninvasive Ventilation for Acute Exacerbations of Chronic Obstructive Pulmonary Disease in the United States, 1998–2008. American Journal of Respiratory and Critical Care Medicine. 2012;185(2):152-9.

13.	Roberts CM, Lopez-Campos JL, Pozo-Rodriguez F, Hartl S. European hospital adherence to GOLD recommendations for chronic obstructive pulmonary disease (COPD) exacerbation admissions. Thorax. 2013;68(12):1169-71.

14.	Confalonieri M, Garuti G, Cattaruzza MS, Osborn JF, Antonelli M, Conti G, Kodric M, Resta O, Marchese S, Gregoretti C, Rossi A, Italian noninvasive positive pressure ventilation study g. A chart of failure risk for noninvasive ventilation in patients with COPD exacerbation. The European respiratory journal. 2005;25(2):348-55.

15.	Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Critical care medicine. 1985;13(10):818-29.

16.	Wildman MJ, Harrison DA, Welch CA, Sanderson C. A new measure of acute physiological derangement for patients with exacerbations of obstructive airways disease: the COPD and Asthma Physiology Score. Respiratory medicine. 2007;101(9):1994-2002.

17.	Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, Lewis SA, Macfarlane JT. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58(5):377-82.

18.	Duan J, Wang S, Liu P, Han X, Tian Y, Gao F, Zhou J, Mou J, Qin Q, Yu J, Bai L, Zhou L, Zhang R. Early prediction of noninvasive ventilation failure in COPD patients: derivation, internal validation, and external validation of a simple risk score. Ann Intensive Care. 2019;9(1):108-.

19.	Steer J, Gibson J, Bourke SC. The DECAF Score: predicting hospital mortality in exacerbations of chronic obstructive pulmonary disease. Thorax. 2012;67(11):970-6.

20.	Echevarria C, Steer J, Heslop-Marshall K, Stenton SC, Hickey PM, Hughes R, Wijesinghe M, Harrison RN, Steen N, Simpson AJ, Gibson GJ, Bourke SC. Validation of the DECAF score to predict hospital mortality in acute exacerbations of COPD. Thorax. 2016;71(2):133-40.

21.	Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55-63.

22.	Field A. Discovering Statistics using IBM SPSS Statistics. 3rd ed: SAGE Publications; 2009.

23.	Bonnett LJ, Snell KIE, Collins GS, Riley RD. Guide to presenting clinical prediction models for use in clinical settings. BMJ. 2019;365:l737.

24.	DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988;44(3):837-45.

25.   Office of National Statistics. Ethnicity Facts and Figures 2018 [Available from: https://www.ethnicity-facts-figures.service.gov.uk/british-population/demographics/age-groups/latest.

26.   Futures R. Dyspnoea, Eosinopenia, Consolidation, Acidaemia and atrial Fibrillation (DECAF) prognostic score 2016 [Available from: https://www.respiratoryfutures.org.uk/features/decaf-prognostic-score/.

27.   Confalonieri M, Potena A, Carbone G, Porta RD, Tolley EA, Umberto Meduri G. Acute respiratory failure in patients with severe community-acquired pneumonia. A prospective randomized evaluation of noninvasive ventilation. Am J Respir Crit Care Med. 1999;160(5 Pt 1):1585-91.

28.   Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. Bmj. 2009;338:b604.

29.   Roberts CM, Stone RA, Buckingham RJ, Pursey NA, Lowe D. Acidosis, non-invasive ventilation and mortality in hospitalised COPD exacerbations. Thorax. 2010.

30.   Jayadev A, Stone R, Steiner MC, McMillan V, Roberts CM. Time to NIV and mortality in AECOPD hospital admissions: an observational study into real world insights from National COPD Audits. BMJ open respiratory research. 2019;6(1):e000444.

Figure 1 The NIVO score.

NIVO score	Points
Chest radiograph consolidation	1
GCS ≤14	1
Atrial Fibrillation	1
pH <7.25	1
Time to Acidaemia >12 hours	2
eMRCD 5a	2
eMRCD 5b	3
	/9

NIVO score: Maximum score of 9 as cannot score for both eMRCD 5a and 5b.

Figure 2 Area under the receiver operated curve and 95% confidence intervals for NIVO score and comparison scores within validation cohort. All scores in 24 hours pre ventilation.
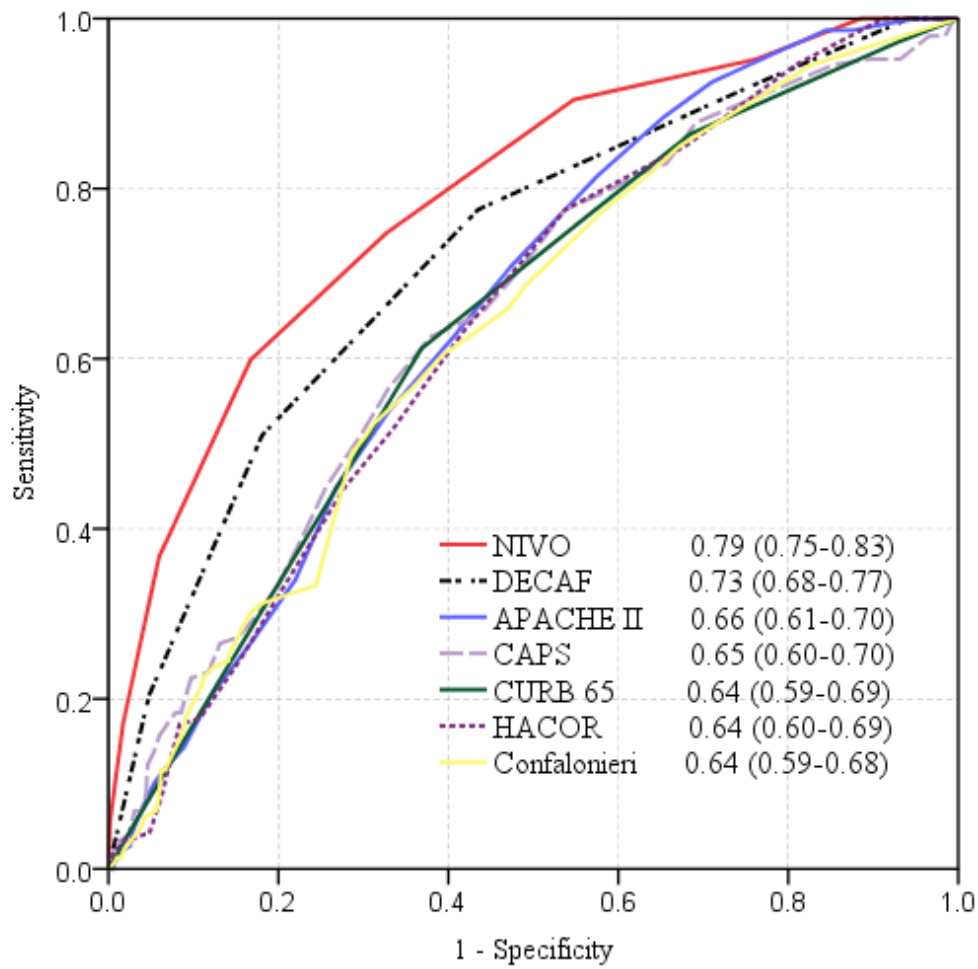
Online Supplement:

Online Supplement I: Missing data for components of NIVO score and comparison scores.

	Derivation			Validation		
Variable	% Missing	Original	Post EM	% Missing	Original	Post EM
pH	0.0	7.24 (0.09)	N/A	0.0	7.24 (0.08)	N/A
Time to acidaemia (min)	0.0	146 (56-852)	N/A	0.0	137 (41-767)	N/A
Temperature (^{o}C)	0.0	35.7 (6.0)	N/A	2.3	36.7 (1.0)	36.7 (1.0)
Systolic BP (mmHg)	1.4	123.9 (31.5)	123.0 (32.4)	2.2	133.4 (34.3)	133.3 (34.0)
Diastolic BP (mmHg)	1.6	70.1 (18.8)	69.7 (19.2)	2.2	75.6 (20.8)	75.4 (20.7)
Heart rate	1.4	112.8 (21.9)	112.4 (22.1)	1.9	110.0 (21.4)	109.9 (21.4)
Respiratory rate	0.0	28.5 (8.2)	N/A	2.2	28.4 (7.4)	28.4 (7.4)
GCS	2.2	13.6 (2.8)	13.5 (2.7)	2.3	13.8 (2.5)	13.8 (2.4)
Sodium (mmol/L)	3.3	136.6 (5.3)	136.6 (5.2)	1.5	136.9 (5.5)	136.9 (5.5)
Potassium (mmol/L))	6.7	4.65 (0.7)	4.63 (0.7)	6.5	4.61 (0.64)	4.61 (0.64)
Urea (mmol/L)	3.3	9.2 (5.7)	9.0 (5.7)	3.1	8.5 (5.3)	8.5 (5.3)
Creatinine (umol/L)	3.3	105.6 (53.3)	105.7 (53.4)	1.8	87.1 (49.4)	87.2 (49.1)
Albumin (g/L)	19.0	37.9 (5.4)	37.9 (5.4)	15.0	37.8 (5.2)	37.7 (5.2)
Haematocrit (L/L)	3.9	0.420 (0.06)	0.420 (0.06)	1.6	0.433 (0.06)	0.433 (0.06)
White cell count (x10^9/L)	3.9	13.7 (6.6)	13.6 (6.6)	1.4	13.0 (6.7)	13.1 (6.8)
Eosinophil Count (x10^9/L)	4.1	0.09 (0.18)	0.10 (0.18)	2.9	0.12 (0.24)	0.12 (0.24)

Online Supplement II: Univariate association with in-hospital mortality (continuous variables) in derivation cohort. Dynamic variables recorded in 24 hours prior to ventilation. Candidate indices for multivariable, regression analysis.

Variable for further evaluation	P Value
Age	<0.001
Atrial Fibrillation present	0.012
Left Ventricular Systolic Dysfunction	0.025
Depression	0.097
eMRCD	<0.001
Long term Oxygen Therapy (LTOT)	0.014
Chest radiograph consolidation	<0.001
Pleural Effusion on admission chest radiograph	<0.001
Confusion present	<0.001
GCS	<0.001
Systolic Blood pressure	0.026
Heart Rate	0.057
Respiratory Rate	0.041
Haemoglobin	<0.001
White Cell Count (WCC)	0.016
Eosinophil Count	<0.001
Urea	<0.001
C- reactive protein (CRP)	0.005
pH	0.003
Base Excess	0.001
Time to acidaemia	<0.001
Albumin **(Not eligible for further exploration due to missing data rate.)**	<0.001

Online Supplement III: Multivariable, forced entry, logistical regression results in validation cohort using 8 categorical variables from derivation cohort. Inpatient mortality as dependent variable.

	B	Significance	(Odds Ratio 95% CI)
Chest radiograph consolidation	0.343	0.106	1.41 (0.93-2.14)
Glasgow Coma Scale (GCS) ≤14	0.650	0.003	1.92 (1.25-2.93)
Atrial fibrillation present	0.838	<0.0001	2.31 (1.45-3.70)
pH <7.25	0.989	<0.0001	2.69 (1.73-4.18)
Time to acidaemia >12 hours	1.300	<0.0001	3.67 (2.35-5.73)
eMRCD 5a	1.378	<0.0001	3.97 (3.34-6.74)
eMRCD 5b	1.892	<0.0001	6.63 (3.73-11.79)
Long term Oxygen Therapy (LTOT)	0.239	0.296	1.27 (0.81-1.99)
Eosinophil count <0.05 x10^9/L	0.107	0.617	1.11 (0.73-1.69)

Intercept -2.976, R^2 0.288, Hosmer and Lemeshow 0.561. Cook's distances and Studentised residuals all within acceptable limits.

Using these eight variables and simple scoring system from derivation (1 point for all except 2 points for 5b) gives AUROC to predict in-hospital mortality of 0.77 (0.73-0.81). I.E an unmodified validation.

Online Supplement IV: the extended medical research council dyspnoea (eMRCD) score and guidance notes.

Extended MRC Dyspnoea (eMRCD) Score **"In the past 3 months, when you were feeling at your best, which of the following statements best describes your level of breathlessness?" (please circle)**	
Only Breathless on strenuous exertion	1
Breathless hurrying on the level or walking up a slight hill	2
Walks slower than contemporaries, or stops when walking on the level for 15 min	3
Stops for breath after walking 100m, or for a few minutes, on the level	4
Too breathless to leave the house unassisted but independent in washing and/ or dressing	5a
Too breathless to leave the house unassisted and requires help with both washing and dressing	5b

Guidance notes:
Remember that you are asking the patient about their level of breathlessness on a good day over the preceding 3 months, not breathlessness during an exacerbation / on admission.

A patient only achieves a higher grade if they are as breathless as defined in that higher grade.
- for example, if worse than defined in eMRCD 3, but not as bad as eMRCD 4, they remain eMRCD 3.

A key distinction is between eMRCD 4 and eMRCD 5a/5b:
- only score 5a or 5b if the patient cannot leave the house without assistance.
- if a patient can only walk 30 to 40 metres, but can leave the house unassisted, they are eMRCD 4.
- if a patient can walk 5 or 10 metres, perhaps from their front door to a car, but need a wheelchair otherwise, they require assistance: eMRCD 5a or 5b. Simple walking aids do not constitute assistance.

If a patient requires assistance in personal washing and dressing they are eMRCD 5b. If they only require assistance in washing or dressing they are eMRCD 5a. Remember to ask about putting on socks and shoes.
If patients are limited for a reason other than breathlessness, score based on their functional limitation.