The Noninvasive Ventilation Outcomes (NIVO) score: prediction of in-hospital mortality in exacerbations of COPD requiring assisted ventilation

Tom Hartley[1,2], Nicholas D. Lane [1,2], John Steer[1,2], Mark W. Elliott[3], Milind P. Sovani[4], Helen Jane Curtis[5], Elizabeth R. Fuller[6,7], Patrick B. Murphy [8,9], Dinesh Shrikrishna[10], Keir E. Lewis[11,12], Neil R. Ward[13], Chris D. Turnbull[14,15], Nicholas Hart [8,9] and Stephen C. Bourke[1,2]

[1]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK. [2]Department of Respiratory Medicine, Northumbria Healthcare NHS Foundation Trust, North Shields, UK. [3]Respiratory Medicine, Leeds Teaching Hospital NHS Trust, Leeds, UK. [4]Department of Respiratory Medicine, Nottingham University Hospitals NHS Trust, Nottingham, UK. [5]Respiratory and Critical Care Medicine, Gateshead Health NHS Foundation Trust, Gateshead, UK. [6]Department of Respiratory Medicine, South Tyneside and Sunderland NHS Foundation Trust, South Shields, UK. [7]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK. [8]Lane Fox Unit, Guy's and St Thomas' NHS Foundation Trust, London, UK. [9]Centre for Human and Applied Physiological Sciences, King's College London, London, UK. [10]Respiratory Medicine, Musgrove Park Hospital, Somerset NHS Foundation Trust, Taunton, UK. [11]Respiratory Medicine, Hywel Dda University Health Board, Llanelli, UK. [12]School of Medicine, University of Swansea, Swansea, UK. [13]Respiratory Medicine, University Hospitals Plymouth NHS Trust, Plymouth, UK. [14]Oxford Centre for Respiratory Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [15]Nuffield Department of Medicine, University of Oxford, Oxford, UK.

Corresponding author: Stephen C. Bourke (Stephen.Bourke@northumbria-healthcare.nhs.uk)

Shareable abstract (@ERSpublications)
The NIVO score was created to predict in-hospital mortality in exacerbations of COPD requiring assisted ventilation. Prospective validation under real-world conditions in 10 UK hospitals shows it easily outperforms existing alternative scores. https://bit.ly/3oKMZdI

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 noninvasive 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 (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) recommendations.

Results Derivation cohort: n=489, in-hospital mortality 25.4%; validation cohort: n=733, in-hospital mortality 20.1%. Using six simple categorised variables (extended Medical Research Council Dyspnoea score 1–4/5a/5b, time from admission to acidaemia >12 h, 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 Noninvasive Ventilation Outcomes (NIVO) score had area under the receiver operating 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 >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 noninvasive ventilation (NIV) [4, 5]. A Cochrane review states that NIV reduces need for intubation by 65% (number needed to treat (NNT) 5) and 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 suggest that fewer than half of all patients with AECOPD and AHRF receive NIV, yet only ∼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 underdeveloped 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 review of national practice in the UK [9]. Such recent, published, national data are 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 nonspecialist 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. [14], 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. In keeping with many predictive models, it has relatively limited external validation. Other predictive tools, such as the APACHE (Acute Physiology and Chronic Health Evaluation) II [15], COPD and Asthma Physiology Score (CAPS) [16], CURB-65 (confusion of new onset, blood urea nitrogen >7 mmol·L^{-1} (19 mg·dL^{-1}), respiratory ⩾30 breaths·min^{-1}; blood pressure <90 mmHg systolic or ⩽60 mmHg diastolic, age ⩾65 years) [17], HACOR (heart rate, acidosis, consciousness, oxygenation, and respiratory rate) [18] or DECAF (Dyspnoea, Eosinopenia, Consolidation, Acidaemia and atrial Fibrillation) score [19, 20] 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 Noninvasive Ventilation Outcomes (NIVO) research group) predicts in-hospital mortality in all exacerbations of COPD whether or not acidaemia is present; the area under the receiver operating characteristic curve (AUROC) is 0.82–0.86 overall, but performance is less strong in patients requiring ventilation.

The absence and 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 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 Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (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 that 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 National Health Service 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 that few patients are denied ventilation on the 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, pre-admission spirometry evidence of airflow obstruction (forced expiratory volume in 1 s (FEV_1)/vital capacity <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 the study or illness other than COPD likely to limit life to <1 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 data, components of comparator models and indices associated with mortality from literature review were collected. Each case record (paper notes and electronic records) was 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 h 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, CAPS, DECAF score, CURB-65) and any index from the derivation study with a univariate association with mortality. Although not pre-specified, the recently developed HACOR score was 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%, ⩾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 algorithm. Data were characterised into mean±SD if parametric, median (interquartile range (IQR)) if nonparametric and as a proportion if categorical. Univariate analysis was performed using the 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. 1) Potential predictor variables were determined: univariate association with mortality (p<0.1); 2) collinearity was assessed; where present, the variable with strongest plausibility or most significant association was used; 3) 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); 4) multivariable analysis; logistical regression with a backward, stepwise entry method was used; 5) the remaining continuous independent predictors in the resultant "full model" were simplified. Dichotomy used the following hierarchical approach: AUROC analysis, results from previous research, a clinically meaningful value or a median split. The extended Medical Research Council dyspnoea scale (eMRCD) has three categories in line with previous research [19, 20]; 6) categorised variables were re-entered into regression analysis to ensure they retained prognostic value. The strongest remaining variables according to their regression coefficient were included and relative weighting was ascribed [23]; 7) 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; 8) discrimination of final model and comparison to other models was by AUROC analysis and statistical difference assessed using the method of DELONG et al. [24]. Performance

of the tool using categorised (simple model) and continuous (full regression equation) indices was compared; 9) 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 were 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 were 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 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 November 2008 and 19 May 2013 to Northumbria Healthcare NHS Foundation Trust. 124 (25.4%) out of 489 died in hospital. Missing data are shown in the supplementary material; table 1 shows population descriptors. Admissions were split between the two Trust sites (52% *versus* 48%). Current smokers represented 48.7%, and total cigarette smoking burden was high: mean±SD 49.5±26.0 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 radiograph consolidation was present in 47.2%.

Time from admission to index episode of acidaemia was 146 (56–852) min with 73.8% occurring in the first 12 h. At ventilation initiation, median inspiratory positive airway pressure/expiratory positive airway pressure was 16/4 cmH_2O, rising to 18/4 cmH_2O at 1 h and 20/4 cmH_2O maximum. Among 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 1 to 3, 21 variables were entered into the regression equation (supplementary material). Univariate significance was unchanged with or without imputed data. Some statistical associations with mortality were rejected: albumin (missing data rates), mean arterial pressure and body mass index (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 4 (regression using continuous variables), 11 variables remained. Step 5 generated categorical variables and multivariable regression was repeated (step 6) (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 base excess associated with mortality). Similarly, in choosing our eight 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 AUROC in the derivation cohort of 0.85 (0.82–0.89).

Validation cohort
Recruitment was open between October 14, 2016 and February 28, 2018, 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; among the eight variables from the derivation project missing data was 0%, except Glasgow Coma Scale (2.3%) and eosinophil count (2.9%), where data were inputted using the expectation-maximisation algorithm (full details in supplementary material).

TABLE 1 Key population descriptors

	Derivation cohort	Validation cohort	Individual validation sites									
			A	B	C	D	E	F	G	H	I	J
Patients 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$_1$ % pred	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
Previous 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 at ventilation	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	10.5±2.8	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
IPAP$_{max}$ 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)

Data are presented as %, mean±SD or median (interquartile range), unless otherwise stated. BMI: body mass index; eMRCD: extended Medical Research Council dyspnoea scale; FEV$_1$: forced expiratory volume in 1 s; LTOT: long-term oxygen therapy; NIV: noninvasive ventilation; HMV: home mechanical ventilation; CO$_2$: arterial carbon dioxide at ventilation; IPAP$_{max}$: maximum achieved inspiratory positive airway pressure; IMV: invasive mechanical ventilation; APACHE: Acute Physiology and Chronic Health Evaluation.

TABLE 2 Derivation cohort: multivariable, logistical regression results using in-hospital mortality as the dependent variable

	B	Significance	OR (95% CI)
Chest radiograph consolidation	1.019	<0.0001	2.77 (1.61–4.76)
Glasgow Coma Scale ⩽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 h	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	0.764	0.012	2.15 (1.19–3.88)
Respiratory rate ⩾30 breaths·min^{-1}	0.675	0.012	1.97 (1.16–3.33)
Eosinophil count <0.05×10^9 cells·L^{-1}	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. eMRCD: extended Medical Research Council dyspnoea scale.

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% medical ward (22.5% mortality), 13.5% high-dependency unit (20.2% mortality) and 7.6% intensive care unit (26.8% mortality). Place of care was not significantly associated with in-hospital mortality. Median (IQR) length of stay was 8 (6–14) days. 98% of participants were Caucasian (95.5% of over-65s in England and Wales are Caucasian) [25]. Significant antecedent tobacco burden was seen, mean±SD 44.8±23.7 pack-years. 68.9% had been admitted to hospital for any reason in the preceding year.

The eight variables from the derivation study were examined (using logistic regression analysis) to see if further simplification was possible (supplementary material). 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 (nonsignificant 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 six-variable model was finalised. Following the removal of variables, re-weighting was applied; of note, re-weighting 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.

Table 4 shows the in-hospital and 90-day mortality by NIVO score and simple risk categories. The NIVO score significantly outperforms pre-specified comparator scores when predicting in-hospital mortality (p<0.001, NIVO versus each other tool, using method described by DELONG et al. [24]). Moreover, if continuous variables were not categorially transformed, the model accuracy is not substantially improved: AUROC 0.80 (0.76–0.84).

TABLE 3 Validation cohort: multivariable, logistical regression results using in-hospital mortality as the dependent variable

	B	Significance	OR (95% CI)
Chest radiograph consolidation	0.358	0.089	1.43 (0.95–2.16)
Glasgow Coma Scale ⩽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 h	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. eMRCD: extended Medical Research Council dyspnoea scale.

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

Discussion

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

NIVO score	Points
Chest radiograph consolidation	1
Glasgow Coma Scale ≤14	1
Atrial fibrillation	1
pH <7.25	1
Time to acidaemia >12 h	2
eMRCD 5a	2
eMRCD 5b	3
Total	9

FIGURE 1 The Noninvasive Ventilation Outcomes (NIVO) score. Maximum score of 9, as cannot score for both extended Medical Research Council dyspnoea scale (eMRCD) 5a and 5b.

NIVO	0.79 (0.75–0.83)
DECAF	0.73 (0.68–0.77)
APACHE II	0.66 (0.61–0.70)
CAPS	0.65 (0.60–0.70)
CURB-65	0.64 (0.59–0.69)
HACOR	0.64 (0.60–0.69)
Confalonieri	0.64 (0.59–0.68)
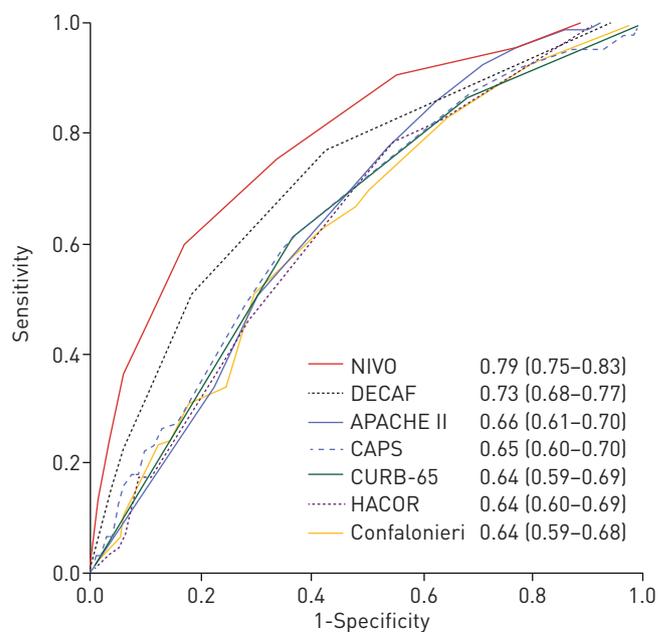
FIGURE 2 Area under the receiver operating characteristic curve (95% CI) for the Noninvasive Ventilation Outcomes (NIVO) score and comparison scores within validation cohort. All scores in 24 h pre-ventilation. DECAF: Dyspnoea, Eosinopenia, Consolidation, Acidaemia and atrial Fibrillation; APACHE: Acute Physiology and Chronic Health Evaluation; CAPS: COPD and Asthma Physiology Score; CURB-65: confusion of new onset, blood urea nitrogen >7 mmol·L^{-1} (19 mg·dL^{-1}), respiratory ⩾30 breaths·min^{-1}; blood pressure <90 mmHg systolic or ⩽60 mmHg diastolic, age ⩾65 years; HACOR: heart rate, acidosis, consciousness, oxygenation and respiratory rate.

TABLE 4 In-hospital and 90-day mortality by Noninvasive Ventilation Outcomes (NIVO) score increment and risk category

	Patients n	In-hospital mortality	90-day mortality
NIVO score			
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

Data are presented as %, unless otherwise stated.

Nearly 40% of patients (270 out of 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 long-term oxygen therapy 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 been treated previously with NIV at the time of their index admission. The inclusion of NIV-naïve patients, those with

TABLE 5 Rules of thumb

	In-hospital mortality		90-day mortality
	Derivation	Validation	Validation
MRCD 5 (eMRCD 5a/5b) + acidaemia >12 h after admission	46/77 (59.7)	50/101 (49.5)	57/101 (56.4)
MRCD 5 (eMRCD 5a/5b) + acidaemia >48 h after admission	27/40 (67.5)	30/45 (66.7)	36/45 (80.0)

Data are presented as n/n (%). eMRCD: extended Medical Research Council dyspnoea scale.

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 radiographic 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 randomised controlled trial 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 probably 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, and 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 six 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 (since the original literature search, the importance of timing has 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 arterial oxygen tension/inspiratory oxygen fraction (F_{iO_2}) ratio. Whether tested in the whole population using estimated values from uncontrolled oxygen or limited to those with known F_{iO_2}, 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 may 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. https://statistics.blf.org.uk/lung-disease-uk-big-picture Date last accessed: 28 August 2020.

2 Royal College of Physicians, British Thoracic Society, British Lung Foundation. Report of The National Chronic Obstructive Pulmonary Disease Audit 2008: Clinical Audit of COPD Exacerbations Admitted to Acute NHS Units Across the UK. 2008. Available from: https://www.rcplondon.ac.uk/projects/outputs/national-copd-audit-2008 Date last accessed: 02 September 2020.

3 Royal college of Physicians, British Thoracic Society. 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. Available from: https://www.rcplondon.ac.uk/projects/outputs/copd-who-cares-matters-clinical-audit-2014 Date last accessed: 02 September 2020.

4 Davidson AC, Banham S, Elliott M, *et al.* 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, *et al.* Official ERS/ATS clinical practice guidelines: noninvasive ventilation for acute respiratory failure. *Eur Respir J* 2017; 50: 1602426.

6   Osadnik CR, Tee VS, Carson-Chahhoud KV, *et al.* Non-invasive ventilation for the management of acute hypercapnic respiratory failure due to exacerbation of chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2017; 7: CD004104.

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. *Lancet* 2000; 355: 1931–1935.

8   Wildman MJ, Sanderson C, Groves J, *et al.* 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: 1132.

9   The National Confidential Enquiry into Patient Outcome and Death. Inspiring Change: a Review of the Quality of Care Provided to Patients Receiving Acute Non-invasive Ventilation. London, Healthcare Quality Improvement Partnership, 2017.

10   Crimi C, Noto A, Princi P, *et al.* A European survey of noninvasive ventilation practices. *Eur Respir J* 2010; 36: 362–369.

11   Bierer GB, Soo Hoo GW. Noninvasive ventilation for acute respiratory failure: a national survey of Veterans Affairs hospitals. *Respir Care* 2009; 54: 1313–1320.

12   Chandra D, Stamm JA, Taylor B, *et al.* Outcomes of noninvasive ventilation for acute exacerbations of chronic obstructive pulmonary disease in the United States, 1998–2008. *Am J Respir Crit Care Med* 2012; 185: 152–159.

13   Roberts CM, Lopez-Campos JL, Pozo-Rodriguez F, *et al.* European hospital adherence to GOLD recommendations for chronic obstructive pulmonary disease (COPD) exacerbation admissions. *Thorax* 2013; 68: 1169–1171.

14   Confalonieri M, Garuti G, Cattaruzza MS, *et al.* A chart of failure risk for noninvasive ventilation in patients with COPD exacerbation. *Eur Respir J* 2005; 25: 348–355.

15   Knaus WA, Draper EA, Wagner DP, *et al.* APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13: 818–829.

16   Wildman MJ, Harrison DA, Welch CA, *et al.* A new measure of acute physiological derangement for patients with exacerbations of obstructive airways disease: the COPD and Asthma Physiology Score. *Respir Med* 2007; 101: 1994–2002.

17   Lim WS, van der Eerden MM, Laing R, *et al.* Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003; 58: 377–382.

18   Duan J, Wang S, Liu P, *et al.* 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: 108.

19   Steer J, Gibson J, Bourke SC. The DECAF Score: predicting hospital mortality in exacerbations of chronic obstructive pulmonary disease. *Thorax* 2012; 67: 970–976.

20   Echevarria C, Steer J, Heslop-Marshall K, *et al.* Validation of the DECAF score to predict hospital mortality in acute exacerbations of COPD. *Thorax* 2016; 71: 133–140.

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

22   Field A. Discovering Statistics using IBM SPSS Statistics. 3rd edn. London, SAGE Publications, 2009.

23   Bonnett LJ, Snell KIE, Collins GS, *et al.* 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: 837–845.

25   Office of National Statistics. Ethnicity Facts and Figures 2018. www.ethnicity-facts-figures.service.gov.uk/ british-population/demographics/age-groups/latest Date last accessed: 02 September 2020. Date last updated: 17 August, 2020.

26   Respiratory Futures. Dyspnoea, Eosinopenia, Consolidation, Acidaemia and atrial Fibrillation (DECAF) prognostic score. www.respiratoryfutures.org.uk/features/decaf-prognostic-score Date last accessed: 2 September 2020. Date last updated: 16 September, 2020.

27   Confalonieri M, Potena A, Carbone G, *et al.* 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: 1585–1591.

28   Royston P, Moons KG, Altman DG, *et al.* Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009; 338: b604.

29   Roberts CM, Stone RA, Buckingham RJ, *et al.* Acidosis, non-invasive ventilation and mortality in hospitalised COPD exacerbations. *Thorax* 2011; 66: 43–48.

30   Jayadev A, Stone R, Steiner MC, *et al.* Time to NIV and mortality in AECOPD hospital admissions: an observational study into real world insights from National COPD Audits. *BMJ Open Respir Res* 2019; 6: e000444.