Prediction of Non-invasive Mechanical Ventilation Response. Moving from Art to Science?

G. Glover¹; S. Fletcher²; A. Esquinas³

¹Guy’s and St Thomas’ NHS Foundation Trust, Critical Care, London, United Kingdom; ²Bradford Teaching Hospitals NHS Foundation Trust, Intensive Care Unit, Bradford, United Kingdom; ³Hospital Morales Meseguer, Intensive Care Unit, Murcia, Spain

Keywords
Non-invasive ventilation (NIV), failure, predictive models, multiclassifiers

Summary
Predicting the outcome from NIV is important and the study by Martin-Gonzalez and colleagues applies data mining techniques to improve our understanding of the field. Nevertheless, the predictor variables must be robust and reliably available before NIV is applied. A predictive model must be generalisable in other clinical settings. Until models such as this are extremely robust in their predictive ability and have been shown to positively influence patient centered outcomes, they may be able to assist decision making but cannot replace clinical judgement by an experienced bedside clinician.

Dear Editor,

We read with interest the article by Martin-Gonzalez and colleagues on the application of data mining techniques to determine the factors that influence success or failure in the use of non-invasive mechanical ventilation (NIV) and to create algorithms to attempt to classify patient outcome [1].

We fully agree with the authors that NIV is a promising therapy which avoids the risks associated with invasive mechanical ventilation (IMV), but which carries its own hazards when used inappropriately [2]. Accurate tools that enable clinicians to correctly judge which patients may benefit, whilst avoiding inappropriate application would be of great value. The authors helpfully describe the data mining methodology, allowing the testing of a greater number of variables than conventional statistical methods, whilst their study benefits from an impressive hospital database. However, we consider that some aspects need to be taken into account for proper clinical extrapolation.

Firstly, if one is to utilise the respiratory failure aetiology as a predictor variable, then it must be recognised that substantial heterogeneity can exist within categories [3]. For example, Thille demonstrated that among 113 patients with acute hypoxic respiratory failure treated with NIV, 82 had acute respiratory distress syndrome (ARDS) and 31 had non-ARDS. Intubation rates significantly differed between ARDS and non-ARDS patients (61% versus 35%, p = 0.015).

Secondarily, the greatest clinical utility of a predictive model is to be able to determine if the treatment is going to be of benefit to the patient, before it is applied. This is especially true in acute care, where the patient's condition may evolve rapidly. Among patients treated with NIV, a significant proportion of patients who fail may do so within the first hour of therapy [4]. In the study by Martin-Gonzalez and colleagues, a number of the selected attributes are only relevant after the initiation of NIV, and hence, the outcome may already have been 'classified' for many patients by the time they are interpretable. For example, four out of five of the most influential attributes determined by the information gain (IG) method will not be known at the time of the initial decision to utilise NIV. Furthermore, the awareness of late factors raises the possibility of self-fulfilment, as clinicians may modify predictors such as fluid balance, with subsequent influence on patient outcome. Nevertheless, many of the factors identified are intuitive, and are also likely to form part of the holistic assessment made by an experienced clinician in evaluating whether to apply NIV and indeed if or when to abandon NIV in favour of escalation to IMV.

Thirdly, a predictive model must be generalisable. The authors studied 389 patients with a total of 410 episodes of NIV, out of a total of 4661 ICU admissions during the study period (just 8.8%). Clinicians in their ICU may have selected patients to be treated with NIV based on patient characteristics, clinician preference and available facilities and it is possible that there were patients subjected to alternative treatments (IMV, extra-corporeal life support, limitation of treatment orders), who may in fact have succeeded with NIV had it been utilised. Thus, there is a potential selection bias, which may limit the generalisability of their findings to all ICU patients.

Because this is a single centre study, the results may be heavily dependent upon the local conditions and may not be applicable in centres with different skill mix, protocols or staffing levels. A predictive model of this nature requires validation in a separate cohort and only then can an analysis determine if it has an ability to positively in-
fluence patient-centered outcomes in a prospective study [5].

Despite these limitations, the results of their work are impressive, with the best classifiers achieving an area under the receiver operator curve of $0.83 \pm 0.06$ and a correct classification of patients of close to 80%. However, these are essentially population statistics and with the risk of misclassification 1:5 or greater, and important risks being associated with NIV when it fails [5], clinicians may rightly be wary about utilising such tools to replace clinical judgement before there is evidence that it will improve patient outcomes [6, 7].

The excellent study by Martin-Gonzalez demonstrates a proof of concept, and how collaboration between technologists and clinicians can help us to answer important clinical questions and ultimately individualise patient treatment, based not only on the best RCT evidence, but also on cumulative real-world experience [1]. We are encouraged to develop collaborative, multicenter networks to develop large databases, designed by clinicians a priori to capture all potentially relevant data, which may enable further refinement of predictive models.

References


Erratum to: “Discussion of ‘Combining Health Data Uses to Ignite Health System Learning’”

S. Denaxas et al.

Unfortunately, the article “Discussion of ‘Combining Health Data Uses to Ignite Health System Learning’” published in Methods of Information in Medicine 2015; 54: 488–499 contained an error on page 488 in the listing of authors following the article’s title. The name of the author H. A. Payne should read T. H. Payne. Please find the corrected list of authors here: S. Denaxas1,2; C. P. Friedman3; A. Geissbuhler4; H. Hemingway1,2; D. Kalra2; M. Kimura6; K. A. Kuhn7; T. H. Payne8; F. G. B. de Quiros9; J. C. Wyatt10,11. Methods Inf Med 2016; 55: 201 http://dx.doi.org/10.3414/ME15-12-0004e