RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

Algorithms to estimate PaCO2 and pH using non invasive parameters for children with Hypoxemic Respiratory Failure

Abstract Words: 285

Word Count Text 3953

Short Title: Non invasive estimates of pH and $PaCO_2$

Robinder G. Khemani MD, MsCI [1,2]
E. Busra Celikkaya, MS [3]
Christian R. Shelton, PhD [3]
Dave Kale, MS [1]
Patrick A. Ross, MD [1,2]
Randall C. Wetzel, MB [1,2]
Christopher J.L. Newth, MD [1,2]

[1] Children's Hospital Los Angeles, Department of Anesthesia and Critical Care
[2] University of Southern California Keck School of Medicine
[3] University of California Riverside, Department of Computer Science

Corresponding Author
Robinder G. Khemani MD, MsCI
Children's Hospital Los Angeles
4650 Sunset Blvd Mailstop 12
Los Angeles, CA 90027
rkhemani@chla.usc.edu
Phone (323) 361-2376; Fax (323) 361-1001

Work performed at Children's Hospital Los Angeles
Presented in abstract form by Dr. Khemani at the American Thoracic Society Meeting, San Francisco, CA May 2012.
Financial support for this study was provided by the Department of Anesthesiology and Critical Care Medicine at Children's Hospital Los Angeles, and the Laura and Leland Whittier Virtual Pediatric Intensive Care Unit
No conflicts of interest for any of the authors.

Literature Search: RGK, PAR, CJLN, EBC
Data Collection: RGK, PAR, CJN
Study Design: RGK, CJN, RCW
Analysis: RGK,EBC, DK, CRS
Manuscript Preparation: RGK, EBC, CS
Manuscript Review: RGK, EBC, PAR, DK, CJN, RCW

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

Abstract

Background: Ventilator management for children with hypoxemic respiratory failure may benefit from ventilator protocols, which rely on blood gases. Accurate non-invasive estimates for pH or $PaCO_2$ could allow frequent ventilator changes to optimize lung protective ventilation strategies. If these models are highly accurate, they can facilitate the development of closed-loop ventilator systems. We sought to develop and test algorithms for estimating pH and $PaCO_2$ from measures of ventilator support, pulse oximetry, and end tidal carbon dioxide ($ETCO_2$). We also sought to determine whether surrogates for changes in dead space can improve prediction.

Methods: Algorithms were developed and tested using 2 datasets from previous published investigations. A baseline model estimated pH and $PaCO_2$ from $ETCO_2$ using the previously observed relationship between $ETCO_2$ and $PaCO_2$ or pH (using Henderson-Hasselbalch equation). We developed a multivariate Gaussian processes (MGP) model incorporating other available non-invasive measurements.

Results: The training dataset had 2,386 observations from 274 children, and the testing dataset 658 observations from 83 children. The baseline model predicted $PaCO_2$ within +/- 7 mmHg of observed $PaCO_2$ 80% of the time. The MGP model improved this to +/- 6 mmHg. When the MGP model predicted $PaCO_2$ between 35-60 mmHg, the 80% prediction interval narrowed to +/- 5 mmHg. For pH, the baseline model predicted pH within +/- 0.07 of observed pH 80% of the time. The MGP model improved this to +/- 0.05.

Conclusions: We have demonstrated a conceptual first step for predictive models that estimate pH and $PaCO_2$ to facilitate clinical decision making for children with lung injury. These models may have some applicability when incorporated in ventilator protocols to encourage practitioners to maintain permissive hypercapnia when using high ventilator support. Refinement with additional data may improve model accuracy.

Abstract Words: 285

Key words: Acute Lung Injury; Pediatrics; Respiration, Artificial; Capnography; Decision Support Techniques

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

Introduction

Ventilator management for children with Acute Lung Injury (ALI) varies widely [1, 2]. Explicit ventilator protocols can standardize mechanical ventilation, provided practitioners follow recommendations [1, 3-6]. Specifically, potentially injurious ventilator settings are frequently not reduced, even with normal or over-ventilated pH or $PaCO_2$.[1] In general, in the acute phase of illness ventilator settings are changed based on arterial blood gases (ABG), requiring an arterial catheter and frequent blood samples, which is challenging in children [7]. Accurate and reliable non-invasive methods to estimate pH or $PaCO_2$ could allow for more frequent ventilator changes during the acute phase of illness to maintain permissive hypercapnia, and help clinical decision making.

Pulse oximetry (SpO_2) is routinely used for clinical decision making [8], and clinicians change PEEP or FiO_2 in response to either PaO_2 or SpO_2, both in the acute phase of illness and during weaning. However, clinicians and ventilator protocols most frequently make decisions to change ventilator rate, tidal volume, or peak inspiratory pressure during the acute phase of illness based on arterial pH or $PaCO_2$. The most widely used non-invasive sensor to estimate adequacy of ventilation is End Tidal Carbon Dioxide ($ETCO_2$). However, the relationship between $ETCO_2$ and $PaCO_2$ changes as a function of alveolar dead space. Additionally, estimating pH from $ETCO_2$ is confounded by changing metabolic acidosis.

At the bedside, one can estimate $PaCO_2$ from $ETCO_2$ using the alveolar dead space fraction ($AVDSF=(PaCO_2-ETCO_2)/PaCO_2$)[9]. While this is not the same as a dead space to tidal volume ratio, which requires volumetric capnography, it is a clinical surrogate [10-13]. Although one can use this value, calculated from simultaneous measurement of $ETCO_2$ and $PaCO_2$ to estimate future $PaCO_2$ from a known value of $ETCO_2$, it will not perform well in the setting of changing alveolar dead space, as may be the case during the acute phase of respiratory illness. To date, most closed-loop algorithms incorporating end tidal CO_2 for ventilator management have been applied to the weaning phase of mechanical ventilation[14].

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

We seek to develop a predictive algorithm to estimate pH and $PaCO_2$ which can account for changing alveolar dead space, for application during the acute phase of illness.

We have previously demonstrated that non-invasive surrogates for intrapulmonary shunt (i.e. Oxygen Saturation Index or SpO_2/FiO_2 Ratio) are correlated with changes in AVDSF[8, 9]. We hypothesize that incorporating non-invasive measures of intrapulmonary shunt, non-invasive continuously available values from the ventilator, and previous known values from the ABG, will permit development of a predictive algorithm to estimate $PaCO_2$ and pH accurately. Such an algorithm could be incorporated into a computer ventilator protocol to encourage lung protective-behavior and permissive hypercapnia for children with lung injury.

Materials and Methods

We developed and tested algorithms using datasets from previous studies on children with acute hypoxemic respiratory failure[1, 8, 9, 15]. We constructed datasets with simultaneous measurements of arterial pH, $PaCO_2$, PaO_2, pulse oximetry (when SpO_2 was $\leq 97\%$), $ETCO_2$ (measured via main stream, with the same adapter size for all children as per ICU standards), and ventilator settings (mode, ventilator rate (VR), peak inspiratory pressure (PIP), positive end expiratory pressure (PEEP), exhaled tidal volume (ml/kg)(V_T), Fraction of Inspired Oxygen (FiO_2)). We created composite variables for deficits in oxygenation including Oxygen Saturation Index (OSI=Mean Airway Pressure * FiO_2*100/SpO_2) and SpO_2/FiO_2 (SF) ratio. We excluded measurements if there was a leak around the endotracheal tube of $\geq 20\%$[16], if it had been > 24 hours since the previous ABG, or if there was only one ABG for an individual patient. The first ABG attained for each patient was used as baseline, and the algorithms generated estimates for pH and $PaCO_2$ at the time of subsequent ABGs. Predicted values for pH and $PaCO_2$ were compared to actual measured values. The study was approved by the Committee on Clinical Investigation at CHLA with a waiver of informed consent (CCI-09-00126 and CCI-09-00287).

Dataset 1: Single Center Dataset

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

We assembled this dataset from a single institution retrospective study. Children (<18 years of age) were included in this study if they were intubated and mechanically ventilated with at least one PF ratio <300 after intubation. Children with left ventricular dysfunction or cyanotic congenital heart disease were excluded. We have previously published the methods regarding data collection, patient characteristics, and ventilator support[1, 15] . We extracted data from the electronic medical record, time ordered per patient. We extracted the closest charted value for SpO_2 and $ETCO_2$ which was at most 1 hour prior to a study ABG. We have previously used this methodology to demonstrate that OSI, AVDSF, and SF ratio correlate with mortality[9].

Dataset 2: Multi-center Dataset

We assembled this dataset from a six-center prospective study in children. Children (<18 years of age) were included in this study if they were intubated and mechanically ventilated with an indwelling arterial line and $SpO_2 \leq 97\%$. Children with left ventricular dysfunction or cyanotic congenital heart disease were excluded. We have previously published the methods regarding data collection, patient characteristics, and ventilator support [8]. SpO_2 and $ETCO_2$ were recorded prospectively precisely at the time of the ABG, with concurrent ventilator settings. Therefore, unlike with the single-center dataset, $ETCO_2$, SPO_2, ventilator settings and ABG results were simultaneous. ABG values were not recorded if the pulse oximetry waveform was inadequate, or if the patient had received endotracheal tube suctioning or invasive procedures for 30 minutes prior to the blood gas. $ETCO_2$ was recorded when available as part of routine care. $ETCO_2$ was not used routinely for all ventilated patients in some of the study ICUs.

Analysis

We report the results of the statistical models trained on dataset 1, and tested on dataset 2. To predict both $PaCO_2$ and pH, we created two models. The first models used the previous simultaneous values for $PaCO_2$ and $ETCO_2$ to calculate AVDSF, which was used to estimate the expected current value for $PaCO_2$, based on a new $ETCO_2$. We used AVDSF instead of the difference between $PaCO_2$ and $ETCO_2$

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

to control for proportionality inaccuracies as $PaCO_2$ increases. The first model for pH uses this estimate for $PaCO_2$ with the serum bicarbonate from the previous ABG to predict pH using the Henderson-Hasselbalch equation.

The second models were generated using a multivariate Gaussian process (MGP): a machine learning technique. A Gaussian process is a probability distribution over a function. The joint values of the function at any subset of times have a multivariate normal distribution, defined by its mean and covariance function. We used a squared exponential (SE) covariance function. The resulting process can be thought of as a generalization of a Bayesian linear regression model applied to higher dimensions. The covariance function of an MGP is represented as a matrix $C(t,t') = cov\left(F(t), F(t')\right) \in \mathbb{R}^{n \times n}$ where the element $C(t,t')_{ij}$ is the correlation between variable i at time t and variable j at time t'. We used a separable model, $C(t,t') = r(t,t')S$ where $r(t,t')$ is the temporal covariance between two time points and $S \in \mathbb{R}^{n \times n}$ is the covariance matrix between the variables. For an offset of observation times $t_1, t_2, \ldots, t_T$ the resulting observations are jointly Gaussian with a covariance matrix of $K \in \mathbb{R}^{nT \times nT}$. We exploit the separable nature of our model and the simultaneity of the observations to avoid explicit computations with such a large matrix. We estimate the covariance matrix S from the training data (dataset 1).

When testing on dataset 2, we assume we know the measurements for all components except for current values of $PaCO_2$ and pH. We predict the mean and covariance of the marginal distributions of $PaCO_2$ and pH at the current time given all known measurements for all the components up until and including the current time.

Additional Models

In addition to the AVDSF model, we computed a model based on minute ventilation which is often calculated at the bedside (estimates the current $PaCO_2$ from the current minute ventilation and previous

$PaCO_2$ and minute ventilation). It did not perform as well as AVDSF, so results are not shown. We tested two additional models to account for changing dead space including hierarchical linear regression, and continuous time based Bayesian network [17], but they were not superior to the MGP model, so the results are not shown. We have previously presented one of these models in abstract form [18].

Outcome Measures

Our primary outcome was the accuracy of different algorithms to predict pH and $PaCO_2$. To evaluate these outcomes we generated 80 and 95% prediction intervals around the point estimate for pH and $PaCO_2$. The purpose of this outcome was to assess whether the model could generate estimates which fall into a range that may be acceptable in certain clinical scenarios 80 or 95% of the time. To mimic the decisions a ventilator protocol would make, we binned observed and predicted values. $PaCO_2$ was binned: <35 mmHg, 35-60mmHg, > 60mmHg. pH was binned using guidance from the ARDS Network protocol: <7.30, 7.30-7.44, ≥ 7.45 [1, 4]. We report percent agreement between observed versus predicted bins, kappa statistics, as well as 80% and 95% prediction intervals for $PaCO_2$ or pH within each of the predicted bins. We also report the percentage of observations which fall within Clinical Laboratory Improvement Amendments (CLIA) standards for $PaCO_2$ (the greater of +/-5mmHg or 8%) and pH (+/-0.04)[19].

Results

Of the 398 children enrolled in the single-center retrospective study (dataset 1), 274 met inclusion criteria with $SpO_2 \leq 97\%$, and $ETCO_2$ results available at most 1 hour prior to the ABG. Of the 103 children without cyanotic congenital heart disease enrolled in the multi-center study, 83 met inclusion criteria with $ETCO_2$ data available at the time of the ABG. Hence, this training dataset had 2,386 observations (aligned $ETCO_2$, SPO_2, ABG, and ventilator data) from 274 children. The testing dataset (dataset 2) had 658 observations from 83 children. In general, the datasets were similar with respect to disease severity, blood gas parameters, and ventilator support. The patients had moderate to severe lung injury, with a

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

median FiO_2 of 0.6 (IQR 0.4,0.8), median PF ratio of 127 (IQR 86,192) and median Oxygenation Index (OI) of 12.6 (IQR 6.7,22) in dataset 1. Results for dataset 2 were similar (Table 1). The median time between observations (ABGs) was 6.5 hours in both datasets (Table 1).

Model Training

The model was trained (parameters were estimated) on dataset 1. Variables included in the model were: $ETCO_2$, OSI, PIP, PEEP, Ventilator Rate, Tidal Volume, Minute Ventilation, an interaction term of $ETCO_2$/Minute Ventilation, Dynamic Compliance of the Respiratory System (Cdyn), and previous values for pH, $PaCO_2$, $ETCO_2$. Other variables considered but not used in the model were: Pressure Support and SF ratio. This is because Pressure Support varied little (the value was almost always 10 cmH_2O) and SF ratio is included in the calculation of OSI which is included in the model.

Prediction of $PaCO_2$ in the testing dataset (Dataset 2)

We constructed a baseline model to predict $PaCO_2$, using the previous AVDSF with the current $ETCO_2$. The predicted values were on average 0.266 mmHg (SD 7.2) higher than the observed values. Overall 67.5% of predicted $PaCO_2$ would fall within CLIA standards (greater of +/-5mmHg or 8% difference) against the measured $PaCO_2$. 80% of the predicted values were within +/- 7 mmHg of observed values and 95% were within +/- 13 mmHg (Figure 1a). When binning the observed and predicted $PaCO_2$, the overall agreement between observed and predicted bins was 89%, with a kappa value of 0.76 (Table 2a). The accuracy was best in the "normal" or "low" $PaCO_2$ bins, where 80% of predicted values were within +/- 6 mmHg of observed values, and 95% within +/- 12 mmHg (Table 3; Figure 2a).

The MGP model derived from dataset 1 performed slightly better than the baseline model using AVDSF. The predicted values were on average 0.02 mmHg (SD 6.1) higher than the observed values. Overall 73.6% of predicted $PaCO_2$ would fall within CLIA standards against the measured $PaCO_2$. For the MGP model, 80% of the predicted values were within +/- 6 mmHg of observed values and 95% were within +/- 11 mmHg (Figure 1b). When binning the observed and predicted $PaCO_2$ the overall agreement between

observed and predicted bins was 91%, with a kappa value of 0.80 (Table 2b). Within each $PaCO_2$ bin, the prediction intervals were narrower than with the AVDSF model, and in the "normal" $PaCO_2$ bin, 80% of predicted values were within +/- 5 mmHg of observed values, and 95% within +/- 10 mmHg. In the "low" $PaCO_2$ bin, 80% of predicted values were within +/- 3 mmHg of observed values, and 95% within +/- 4 mmHg (Table 3; Figure 2b).

Prediction of pH in the testing dataset (Dataset 2)

We used predicted $PaCO_2$ from the AVDSF equation with the calculated serum bicarbonate (HCO_3) from the previous ABG with the Henderson-Hasselbalch equation to predict pH. The predicted values were on average 0.004 (SD 0.064) lower than the observed values. Overall 59.3% of predicted pH would fall within CLIA standards (+/-0.04) against the measured pH. Using this model, 80% of the predicted pH values were within +/- 0.07 of observed values and 95% were within +/- 0.13 (Figure 3a). When binning the observed and predicted pH, the overall agreement between observed and predicted bins was 70%, with a kappa value of 0.48 (Table 4a). The best accuracy was in the "normal" pH bin, where 80% of the predicted values were within +/- 0.06 of observed values and 95% within +/- 0.10 (Table 3; Figure 4a).

The same MGP model (trained on dataset 1) was used to predict pH on dataset 2. In general, the MGP model was slightly superior to the model using the AVDSF and Henderson-Hasselbalch equation. The predicted values were on average 0.002 (SD 0.05) lower than the observed values. Overall 67.6% of predicted pH would fall within CLIA standards against the measured pH. Using the MGP model, 80% of the predicted pH values were within +/- 0.05 of observed values and 95% were within +/- 0.10 (Figure 3b). When binning the observed and predicted pH, the overall agreement between observed and predicted bins was 72%, with a kappa value of 0.49 (Table 4b). Within each pH bin, the prediction intervals are narrower than the Henderson-Hasselbalch model, and in the "normal" pH bin, 80% of the predicted values were within +/- 0.05 of observed values and 95% within +/- 0.10. In the "high" pH bin, 80% of the predicted values were within +/- 0.05 of observed values and 95% within +/- 0.07 (Table 3; Figure 4b).

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

Alternate Training and Testing Datasets

We repeated the analysis using dataset 2 as the training dataset, and dataset 1 as the testing dataset. In general, each model had larger prediction intervals and less agreement between observed and predicted bins of pH and $PaCO_2$ (analysis not shown).

Discussion

We have demonstrated a first step in the application of machine learning algorithms to estimate pH and $PaCO_2$ to facilitate decision making regarding ventilator management for children with moderate to severe lung injury. Over the entire range of predicted values, these models may not yet have an acceptable level of accuracy to replace blood gas sampling. However, these algorithms may be useful in certain clinical scenarios such as decreasing potentially injurious ventilator settings for children who fall in the "over-ventilated range" through a protocol with standardized ventilator decisions. With model refinement, it may become more clinically acceptable in other scenarios. Furthermore, while models which use previous known relationships between $ETCO_2$ and $PaCO_2$ perform reasonably, we can modestly improve the accuracy by incorporating non-invasive markers of oxygenation as well as changes to ventilator settings. It may be that these are surrogates for changes in alveolar dead space. Our multivariate Gaussian process model predicts $PaCO_2$ with 80% prediction intervals of +/-6 mmHg and pH with 80% prediction intervals of +/- 0.05. It performed best in middle or over-ventilated range of pH and $PaCO_2$. For example, if the model predicted $PaCO_2$ to be 50 mmHg, 80% of the time the actual $PaCO_2$ (if one were to draw an ABG) would be between 45 and 55 mmHg, and 95% of the time it would be between 40 and 60 mmHg. For pH in the over-ventilated range, for example, if the model predicted pH to be 7.45, 80% of the time the actual pH would be between 7.4 and 7.5, and 95% of the time it would be between 7.38 and 7.52. Overall, the model would predict $PaCO_2$ with CLIA acceptable equivalence to a blood gas machine 74% of the time and pH 67% of the time.

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

Some may believe that the confidence in these predictions is not adequate for clinical decision making. While in medicine we strive for 95% certainty for statistical significance, clinical decisions are often based on much more uncertainty than 20%. For example, pulse oximetry, in a low range (< 87%) may have 95% prediction intervals greater than +/- 10% against co-oximetry.[20] For example, if the pulse oximeter is reading 85%, then 50% of the time the actual SaO_2 on co-oximetry would lie between 75 and 83%, but for 95% certainty, the SaO_2 could range from 64 to 89%. Nevertheless, pulse oximetry in a low range is routinely used for clinical decision making for children with cyanotic congenital heart disease. While SpO_2 is more accurate in the range frequently seen for children with ALI, this example is meant to illustrate that parameters practitioners routinely use for clinical decisions relating to mechanical ventilation may have more uncertainty than 20%. As such, 80% certainty that the $PaCO_2$ is in a range of 10 mmHg or pH in a range of 0.1 (as seen in our model in the "normal or overventilated range") may be acceptable in certain situations to facilitate a clinical decision such as decreasing potentially injurious ventilator settings when the value is predicted to be normal or high, embracing a permissive hypercapnia strategy for ALI.

We believe there are several potential applications of these algorithms. First, non-invasive estimates of $PaCO_2$ or pH may decrease the number of ABGs. If the clinician can be 80% confident that the $PaCO_2$ lies within a range of 10 mmHg, or the pH within a range of 0.1, he may be willing to forgo an ABG, and instead change the ventilator. This is likely applicable in the "over-ventilated" range for patients with ALI, encouraging more continual lung protective behavior. Leaving the decision to act on the estimate open-loop allows the provider flexibility regarding their comfort with the reported level of certainty, as there may be scenarios when this level of accuracy is not acceptable.

Second, these continuously available estimates may facilitate standardized assessment of ventilator support and adherence to ventilator protocols. For example, an open-loop computer protocol could be developed which requires an assessment every 2 - 4 hours. The decision support tool could display an estimate for the predicted pH, with a prediction interval. The clinician could accept or reject the

protocol's recommendation, or obtain an ABG, if he is uncomfortable with the potential error in the prediction at that time point. As we see in this analysis, the majority of blood gas values for children with ALI lie in a normal range, where the model performs well (> 90% agreement in $PaCO_2$ bins). We have previously demonstrated there are many lost opportunities to be lung protective, when clinicians don't wean the ventilator, even when settings are high and the pH is normal or high [1]. A continuously available estimate of pH or $PaCO_2$ may make clinicians more willing to wean ventilators, because they can more closely monitor the effects of their change. However, as is clear from the analysis, these algorithms have limitations, particularly in the low pH ranges, where there is substantially more uncertainty in the actual pH, seen in the limits of agreement as well as in the Bland-Altman Plots.

We specifically developed this model using data from children with lung injury. We felt it important to start with this population because it represents a more difficult scenario: as children with lung injury have dynamic and changing degrees of dead space. It is likely that both models which assume no change in dead space (such as the AVDSF model) and those that try to capture surrogates for changing dead space (i.e. MGP model) would perform even better in children with minimal or no lung injury. This should, of course, be tested. The median AVDSF in these two datasets were about 0.23, in line with our previous publication on AVDSF which demonstrates that such values are independently associated with mortality[9]. As such, these datasets represent the lung-injured children we often take care of in our intensive care units, in whom lung protective ventilation has the potential to improve outcomes.

While the predictive ability of the MGP algorithm is fair, it is not ready for a closed loop system (where the provider's feedback is not required to change the ventilator), and will not replace blood gases. In fact, these models are reliant upon blood gasses for their development and calibration, and are meant to facilitate decision making at times between blood gasses. This analysis was meant as a first step, the model must be refined with additional data and tighter prediction intervals, before the loop on ventilation is ready to be closed.

To our knowledge, this is the first application of Gaussian processes, a machine learning technique, to predict pH or $PaCO_2$. However, machine learning techniques have been used extensively in medicine [21] for gene expression studies [22-24], classification of cardiac arrhythmias [25], predicting morbidity after coronary artery bypass surgery [26], and predicting when weaning from ventilator support should begin [27]. Gaussian processes have been applied in adults with ALI to model the pressure-volume curve to titrate PEEP [28]. These techniques have been used in industries outside of medicine for years and while the methods may appear complicated, the algorithms are not computationally challenging. Therefore, they can easily be applied in most ICUs, using basic computers.

There are limitations to our analysis. First, this represents secondary analysis of data, which is inherently limited in reliability and accuracy. Second, the algorithms performed worse when trained on the multi-center dataset and tested on the single center dataset. We believe this is a function of the simultaneous assessments of SPO_2, $ETCO_2$ and ABG values in the prospective multi-center dataset, as compared to the retrospective single-center dataset. Because the real-time application of this will be with controlled, continuously available data, we believe that the algorithm will more likely perform as it did with the multi-center data. This of course needs to be tested. Third, we elected not to split the multi-center dataset into training and testing data, because we were worried about sample size. Fourth, the testing dataset was relatively small (84 patients), and the algorithm should be evaluated in another group of patients. Fifth, there was no hemodynamic data, which will also affect changes in dead space. Sixth, the MGP model had a more visible proportional bias on Bland Altman analysis than AVDSF, and may perform differently when applied to a different validation dataset. This should be tested. Seventh, although application of such an algorithm may have an intention to reduce the frequency of blood gasses, it is possible that the early phases of deployment of such an algorithm may prompt clinicians to get more blood gases, to verify what the algorithm is displaying. In all likelihood such a phenomenon would be transient, and once clinicians became more comfortable with its accuracy, they would draw fewer blood gases.

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

We may be able to improve model accuracy with temporally continuous values for $ETCO_2$ and SpO_2, but this may require alternative analytic methodologies to reduce the dimensionality of the data. We may also be able to improve the accuracy by incorporating hemodynamic variables such as heart rate and blood pressure. Finally, these algorithms will likely perform better in less acute phases of mechanical ventilation, such as weaning, when dead space changes less frequently. These hypotheses need to be tested.

Conclusions

Continuously available, non-invasive measurements and ventilator settings may be helpful for predictive algorithms which estimate $PaCO_2$ and pH for children with hypoxemic respiratory failure. The current level of accuracy offers some applications, particularly for standardizing decisions for decreasing ventilator support in line with lung protective strategies when the pH or $PaCO_2$ is predicted to be normal or over-ventilated. With continued model refinement, it is possible that these algorithms can be used for decision support by incorporating them into computer ventilator protocols. These algorithms should be refined and tested with additional prospectively gathered data from mechanically ventilated children with a wide range of severity of lung injury and hemodynamic support.

Acknowledgements

The authors would like to acknowledge the investigators and research personnel from the multi-center study on pulse-oximetry based markers of lung disease severity for their contributions to the data collection for this analysis.

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

References

1. Khemani RG, Sward K, Morris A, Dean JM, Newth CJ, on behalf of the Collaborative Pediatric Crticial Care Research Network (CPCCRN). Variability in usual care mechanical ventilation for pediatric acute lung injury: the potential benefit of a lung protective computer protocol. Intensive Care Med 2011;37(11):1840-1848.
2. Santschi M, Jouvet P, Leclerc F, Gauvin F, Newth CJ, Carroll C, et al. Acute lung injury in children:therapeutic practice and feasibility of international clinical trials. Pediatr Crit Care Med 2010;11(6):681-689.
3. Fan E, Needham DM, Stewart TE, Fan E, Needham DM, Stewart TE. Ventilatory management of acute lung injury and acute respiratory distress syndrome. JAMA 2005;294(22):2889-2896.
4. ARDS Network (ARDSnet). Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 2000;342(18):1301-1308.
5. Putensen C, Theuerkauf N, Zinserling J, Wrigge H, Pelosi P. Meta-analysis: ventilation strategies and outcomes of the acute respiratory distress syndrome and acute lung injury. Ann Intern Med 2009;151(8):566-576.
6. Khemani RG, Newth CJL. The design of future pediatric mechanical ventilation trials for acute lung injury. Am J Respir Crit Care Med 2010;182(12):1465-1474.
7. Khemani RG, Markovitz BP, Curley MAQ. Characteristics of children intubated and mechanically ventilated in 16 PICUs. Chest 2009;136(3):765-771.
8. Khemani RG, Thomas NJ, Venkatachalam V, Scimeme JP, Berutti T, Schneider JB, et al. Comparison of SpO2 to PaO2 based markers of lung disease severity for children with acute lung injury. Critical Care Medicine 2012;40(4):1309-1316.
9. Ghuman AK, Newth CJ, Khemani RG. The association between the end tidal alveolar dead space fraction and mortality in pediatric acute hypoxemic respiratory failure. Pediatr Crit Care Med 2012;13(1):11-15.
10. Frankenfield DC, Alam S, Bekteshi E, Vender RL, Frankenfield DC, Alam S, et al. Predicting dead space ventilation in critically ill patients using clinically available data. Critical Care Medicine 2010;38(1):288-291.
11. Koulouris NG, Latsi P, Dimitroulis J, Jordanoglou B, Gaga M, Jordanoglou J. Noninvasive measurement of mean alveolar carbon dioxide tension and Bohr's dead space during tidal breathing. Eur Respir J 2001;17(6):1167-1174.
12. Hardman JG, Aitkenhead AR, Hardman JG, Aitkenhead AR. Estimating alveolar dead space from the arterial to end-tidal CO(2) gradient: a modeling analysis. Anesth Analg 2003;97(6):1846-1851.
13. McSwain SD, Hamel DS, Smith PB, Gentile MA, Srinivasan S, Meliones JN, et al. End-tidal and arterial carbon dioxide measurements correlate across all levels of physiologic dead space. Respir Care 2010;55(3):288-293.
14. Jouvet PA, Payen V, Gauvin F, Emeriaud G, Lacroix J, Jouvet PA, et al. Weaning children from mechanical ventilation with a computer-driven protocol: a pilot trial. Intensive Care Med 2013;39(5):919-925.

15. Khemani RG, Conti D, Alonzo TA, Bart RD, Newth CJ. Effect of tidal volume in children with acute hypoxemic respiratory failure. Intensive Care Med 2009;35(8):1428-1437.
16. Main E, Castle R, Stocks J, James I, Hatch D. The influence of endotracheal tube leak on the assessment of respiratory function in ventilated children. Intensive Care Med 2001;27(11):1788-1797.
17. Nodelman U, Shelton CR, Koller D. Continuous Time Bayesian Networks. Proceedings of the Eighteenth International Conference of Uncertainty in Artificial Intelligence 2002:378-387.
18. Khemani R, Kale D, Ross P, Wetzel R, Newth C. Predicting Serum pH using non-invasive measurements to enable decision support for mechancially ventialted children with acute lung injury. American Thoracic Society International Conference 2012;A105:28307.
19. CLIA. CLIA proficiency testing criteria for acceptable analytic performance. Federal Register 1992;57(40):7002-7186.
20. Ross PA, Newth CJL, Khemani RG. Accuracy of Pulse Oximetry in Children. Pediatrics 2014;133(1):22-30.
21. Meyfroidt G, Guiza F, Ramon J, Bruynooghe M, Meyfroidt G, Guiza F, et al. Machine learning techniques to examine large patient databases. Best Pract Res Clin Anaesthesiol 2009;23(1):127-143.
22. Statnikov A, Aliferis CF, Tsamardinos I, Hardin D, Levy S, Statnikov A, et al. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. Bioinformatics 2005;21(5):631-643.
23. Zhao X, Cheung LW, Zhao X, Cheung LW-K. Kernel-imbedded Gaussian processes for disease classification using microarray gene expression data. BMC Bioinformatics 2007;8:67.
24. Pirooznia M, Yang JY, Yang MQ, Deng Y, Pirooznia M, Yang JY, et al. A comparative study of different machine learning methods on microarray gene expression data. BMC Genomics 2008;9 (Suppl 1):S13.
25. Asl BM, Setarehdan SK, Mohebbi M, Asl BM, Setarehdan SK, Mohebbi M. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. Artificial Intelligence in Medicine 2008;44(1):51-64.
26. Biagioli B, Scolletta S, Cevenini G, Barbini E, Giomarelli P, Barbini P. A multivariate Bayesian model for assessing morbidity after coronary artery surgery. Critical care (London, England) 2006;10(3):R94.
27. Blockeel H, Bruynooghe M, Loon KV, Aerts M, Berckmans D, Meyfroidt G, et al. Time-series analysis techniques combined with Gaussian Process Classifiers for prediction of clinical stability after coronary bypass surgery. Proceedings of the Sixth IASTED International Conference on Biomedical Engineering 2008:216-221.
28. Ganzert S, Kramer S, Moller K, Steinmann D, Guttmann J. Prediction of Mechanical Lung Parameters Using Gaussian Process Models. Artificial Intelligence in Medicine 2009;Proceedings of the 12th Conference on Artificial Intelligence in Medicine:380-384.

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

Figure Legends

Figure 1a: Observed versus predicted values of $PaCO_2$ based on the AVDSF equation to estimate $PaCO_2$ from the current value of $ETCO_2$, knowing the previous relationship between $PaCO_2$ and $ETCO_2$. Red bounds are the 80% prediction intervals of +/- 7 mmHg and green bounds are the 95% prediction intervals of +/- 13 mmHg. Shaded boxes represent the low, normal, and high bins. The points which lie in the shaded boxes would be classified into the correct bin (89% agreement, kappa = 0.76).

Figure 1b: Observed versus predicted values of $PaCO_2$ based on the MGP model to estimate $PaCO_2$ from the $ETCO_2$, ventilator support, and previous known values of pH and $PaCO_2$. Red bounds are the 80% prediction intervals of +/- 6 mmHg and green bounds are the 95% prediction intervals of +/- 11 mmHg. Shaded boxes represent the low, normal, and high bins. The points which lie in the shaded boxes would be classified into the correct bins (91% agreement, kappa = 0.80).

Figure 2a: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each $PaCO_2$ bin for the AVDSF model. For $PaCO_2$ <35 mean bias was -2.69 (95% Limits of Agreement (LA) +/-11.8); for $PaCO_2$ between 35-60 mean bias was -0.5 (95%LA +/-11.1); for $PaCO_2$ >60 mean bias was 0.6 (95% LA +/-19.5).

Figure 2b: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each $PaCO_2$ bin for the MGP model. For $PaCO_2$ <35 mean bias was -4.0 (95% Limits of Agreement (LA) +/- 9.0); for $PaCO_2$ between 35-60 mean bias was -0.8 (95%LA +/-8.7); for $PaCO_2$ >60 mean bias was 2.1 (95% LA +/-16.6).

Figure 3a: Observed versus predicted values of pH based on the AVDSF equation to estimate $PaCO_2$, and using the previous calculated bicarbonate value to plug into the Henderson-Hasselbalch equation. Red bounds are the 80% prediction intervals of +/- 0.07 and green bounds are the 95% prediction intervals of +/- 0.13. Shaded boxes represent the low, normal, and high bins. The points which lie in the shaded boxes would be classified into the correct bins (70% agreement, kappa = 0.48).

Figure 3b: Observed versus predicted values of pH based on the MGP model to estimate pH from the $ETCO_2$, ventilator support, and previous known values of pH and $PaCO_2$. Red bounds are the 80% prediction intervals of +/- 0.05 and green bounds are the 95% prediction intervals of +/- 0.10. Shaded boxes represent the low, normal, and high bins. The points which lie in the shaded boxes would be classified into the correct bins (72% agreement, kappa = 0.49).

Figure 4a: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each pH bin based using the AVDSF equation to estimate $PaCO_2$, and the previous calculated bicarbonate value to plug into the Henderson-Hasselbalch equation to estimate pH. For pH <7.3 mean bias was -0.017 (95% Limits of Agreement (LA) +/-0.17); for pH between 7.3-7.45 mean bias was 0.001 (95%LA +/-0.115); for pH≥ 7.45 mean bias was 0.015 (95% LA +/-0.113).

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

Figure 4b: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each pH bin for the MGP model. For pH <7.3 mean bias was -0.059 (95% Limits of Agreement (LA) +/-0.13); for pH between 7.3-7.45 mean bias was 0.001 (95%LA +/-0.078); for pH≥ 7.45 mean bias was 0.032 (95% LA +/-0.072).

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

	Training Dataset	Testing Dataset
Patients (n)	274	83
Observations (n)	2386	658
Observations per Patient	5 (2,11)	4 (2,8)
Age (mo)	17 (8,36)	36 (5,134)
Weight (kg)	4.6 (1.1,11.6)	13.5 (5.0,35.9)
Gender (Female)	115 (42%)	27 (33%)
Arterial Blood Gas		
pH	7.37 (7.30,7.44)	7.42 (7.35,7.46)
$PaCO_2$	49 (41,60)	51 (44,63)
PaO_2	72 (61,88)	69 (61,81)
Time between ABGs (hrs)	6.4 (3.6, 11.7)	6.5 (2.0,12.4)
Non-Invasive Support		
SpO_2	95 (92,97)	94 (92,96)
$ETCO_2$	38 (31,45)	40 (35,47)
Time between $ETCO_2$ and ABG (min)	28 (20,46)	Simultaneous
Ventilator Settings		
PIP	30 (25, 35)	30 (26,37)
PEEP	10 (6,12)	8 (5,10)
Mean Airway Pressure	16 (12,20)	16 (13,20)
FiO_2	0.60 (0.41,0.80)	0.55 (0.43,0.63)
V_TE (ml/kg)	7.2 (5.8,8.8)	9 (7,11)
Ventilator Rate	20 (16,26)	18 (16,24)
Lung Disease Severity		
Oxygen Saturation Index (OSI)	10 (6, 16)	9.5 (6.2, 14)
Oxygenation Index (OI)	12.6 (6.7,22.0)	12.6 (7.7,18.8)
SpO_2/FiO_2 (SF) Ratio	162 (119,218)	166 (145,216)
PaO_2/FiO_2 (PF) Ratio	127 (86,192)	130 (95, 178)
Alveolar Dead Space Fraction $(PaCO_2\text{-}ETCO_2)/PaCO_2$	0.24 (0.14,0.34)	0.22 (0.15,0.30)

Table 1: Descriptive statistics of Training and Testing Datasets. Data is presented as median and inter-quartile range, unless otherwise specified.

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

a

Predicted $PaCO_2$	Actual $PaCO_2$			Total
	< 35	**35-60**	**>60**	
<35	13 (52%)	12 (48%)	0	25 (4%)
35-60	5 (1%)	402 (93%)	26 (6%)	433 (66%)
>60	0	30 (15%)	170 (85%)	200 (30%)
Total	18	444	196	658

b

Predicted $PaCO_2$	Actual $PaCO_2$			Total
	< 35	**35-60**	**>60**	
<35	8 (73%)	3 (27%)	0	11 (2%)
35-60	10 (2%)	426 (92%)	30 (6%)	466 (71%)
>60	0	15 (8%)	166 (92%)	181 (30%)
Total	18	444	196	658

Table 2 (a, top) and (b, bottom): Observed versus predicted bins of $PaCO_2$ generated from the AVDSF model for $PaCO_2$ (a, top) and the MGP model for $PaCO_2$ (b, bottom). Overall agreement for the AVDSF model was 89% with a kappa of 0.76, compared to 91% with a kappa of 0.80 for the MGP model. The percentages in each cell are calculated across the rows, to represent the percentage of predicted values in which the actual $PaCO_2$ fell in each bin. For example, for the AVDSF model, $PaCO_2$ was predicted to be between 35 and 60 mmHg 433 times and 402 (93%) of these times the actual $PaCO_2$ was also between 35 and 60 mmHg. The MGP model has better agreement between observed and predicted $PaCO_2$ bins than the AVDSF model, particularly when $PacO_2$ is estimated to be < 35 or > 60 mmHg.

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

	AVDSF Model		MGP Model	
Predicted $PaCO_2$	**80% PI**	**95% PI**	**80% PI**	**95% PI**
<35 mmHg	6	12	3	4
35-60 mmHg	6	12	5	10
>60 mmHg	9	17	8	15
	Henderson-Hasselbalch Model		**MGP model**	
Predicted pH	**80% PI**	**95% PI**	**80% PI**	**95% PI**
< 7.30	0.12	0.24	0.10	0.15
7.30-7.44	0.06	0.10	0.05	0.10
≥ 7.45	0.06	0.12	0.05	0.07

Table 3: The 80 and 95% prediction intervals for the AVDSF model and MGP model to predict $PaCO_2$ and the AVDSF model using Henderson-Hasselbalch and MGP model to predict pH. For the AVDSF model, when $PaCO_2$ is between 35-60 mmHg, 80% of predicted values for $PaCO_2$ would fall within + or – 6 mmHg of the observed values. For the MGP model, the prediction intervals are narrower than the AVDSF model in all 3 ranges of predicted $PaCO_2$. For the AVDSF model using Henderson-Hasselbalch, when pH is between 7.30-7.44, 80% of predicted values for pH would fall within + or – 0.06 of the observed values. For the MGP model, the prediction intervals are narrower than the Henderson-Hasselbalch model in all 3 ranges of predicted pH.

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

Predicted pH	Actual pH			Total
	< 7.3	7.3-7.44	≥7.45	
< 7.3	58 (60%)	36 (38%)	2 (2%)	96 (14%)
7.3-7.44	20 (6%)	264 (73%)	76 (21%)	360 (55%)
≥7.45	2 (1%)	63 (31%)	137 (68%)	202 (31%)
Total	80	363	215	658

Predicted pH	Actual pH			Total
	< 7.3	7.3-7.44	≥7.45	
< 7.3	49 (77%)	15 (23%)	0 (0%)	64 (10%)
7.3-7.44	30 (7%)	305 (71%)	93 (22%)	428 (65%)
≥7.45	1 (1%)	43 (26%)	122 (73%)	166 (25%)
Total	80	363	215	658

Table 4 (a, top) and (b, bottom): Observed versus predicted bins of pH generated from the AVDSF model using Henderson-Hasselbalch for pH (a, top) and the MGP model for pH (b, bottom). Overall agreement for the AVDSF model using Henderson-Hasselbalch was 70% with a kappa of 0.48, compared to 72% with a kappa of 0.49 for the MGP model. The percentages in each cell are calculated across the rows, to represent the percentage of predicted values in which the actual pH fell in each bin. For example, from the AVDSF model using Henderson-Hasselbalch, pH was predicted to be between 7.3 and 7.44 360 times and 264 (73%) of these times the actual pH was also between 7.3 and 7.44. The MGP model has better agreement between observed and predicted pH bins than the AVDSF model, particularly when pH is estimated to be < 7.3 or > 7.45.

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

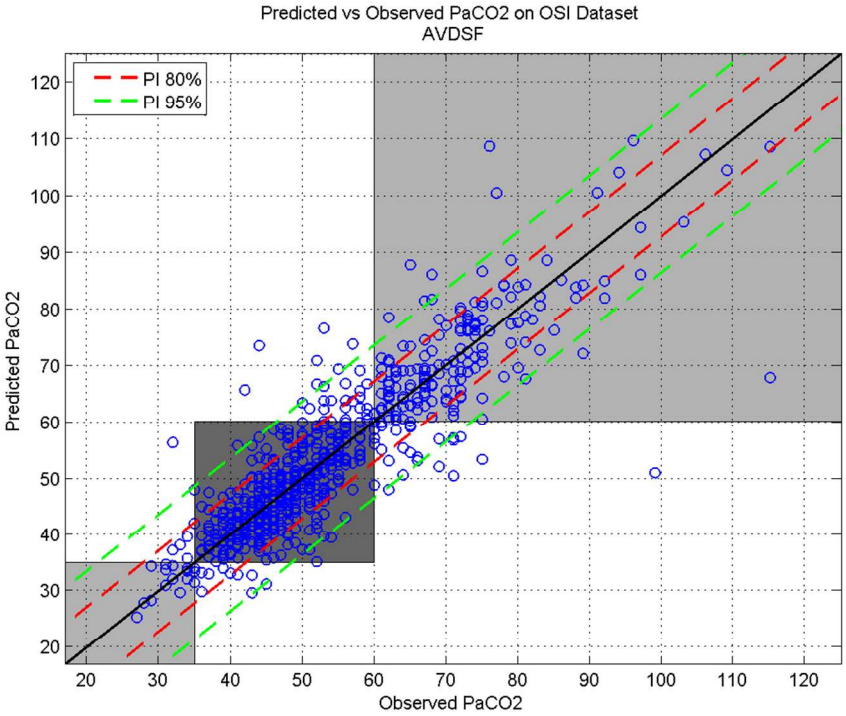

Figure 1a: Observed versus predicted values of PaCO2 based on the AVDSF equation to estimate PaCO2 from the current value of ETCO2, knowing the previous relationship between PaCO2 and ETCO2. Red bounds are the 80% prediction intervals of +/- 7 mmHg and green bounds are the 95% prediction intervals of +/- 13 mmHg. Shaded boxes represent the low, normal, and high bins. The points which lie in the shaded boxes would be classified into the correct bin (89% agreement, kappa = 0.76).
203x152mm (150 x 150 DPI)

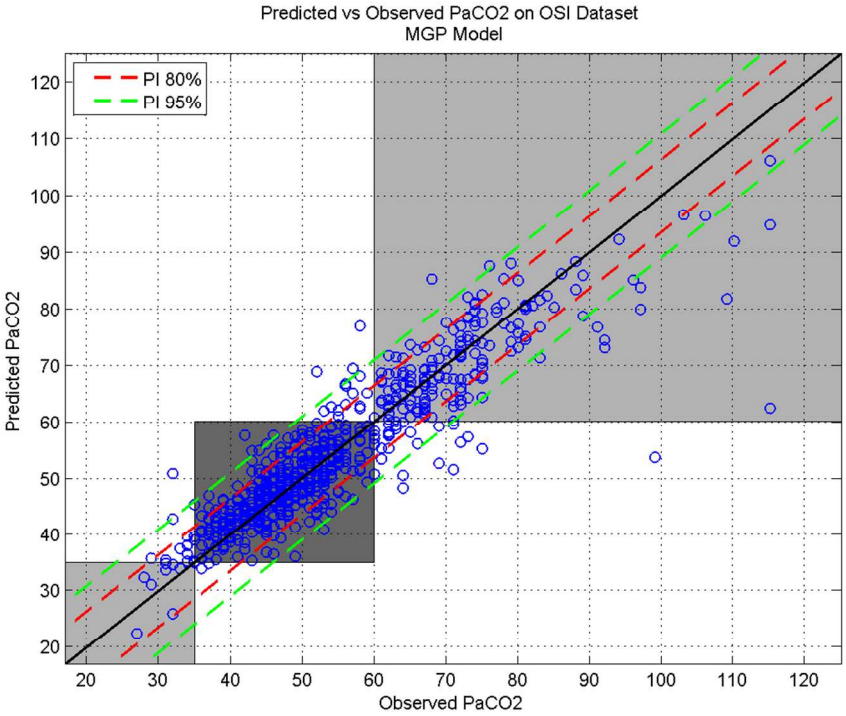

Figure 1b: Observed versus predicted values of PaCO2 based on the MGP model to estimate PaCO2 from the ETCO2, ventilator support, and previous known values of pH and PaCO2. Red bounds are the 80% prediction intervals of +/- 6 mmHg and green bounds are the 95% prediction intervals of +/- 11 mmHg. Shaded boxes represent the low, normal, and high bins. The points which lie in the shaded boxes would be classified into the correct bins (91% agreement, kappa = 0.80).
203x152mm (150 x 150 DPI)

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

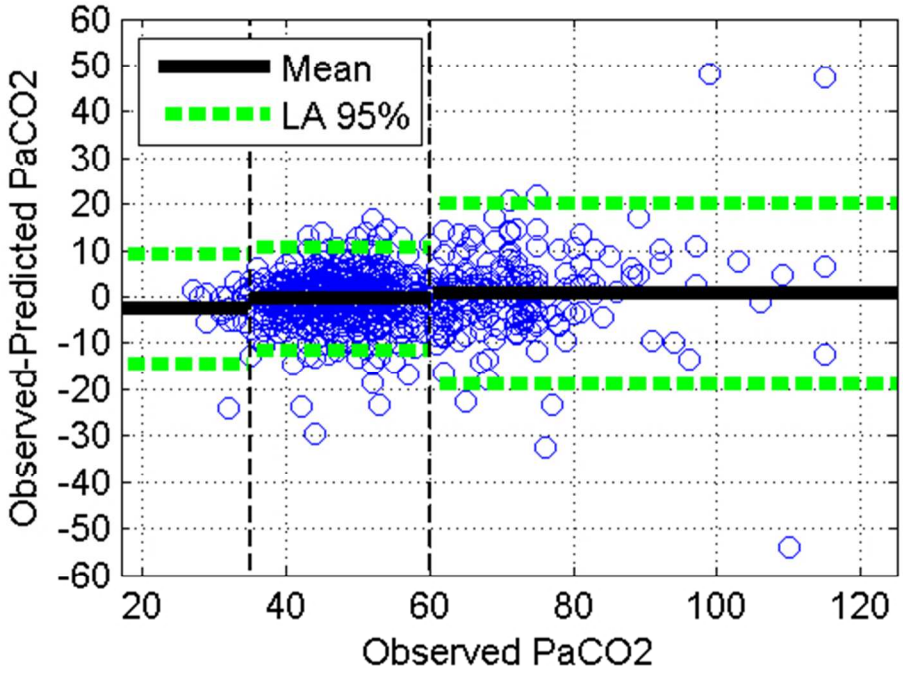

Figure 2a: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each PaCO2 bin for the AVDSF model. For PaCO2 <35 mean bias was -2.69 (95% Limits of Agreement (LA) +/-11.8); for PaCO2 between 35-60 mean bias was -0.5 (95%LA +/-11.1); for PaCO2 >60 mean bias was 0.6 (95% LA +/-19.5).
100x71mm (150 x 150 DPI)

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

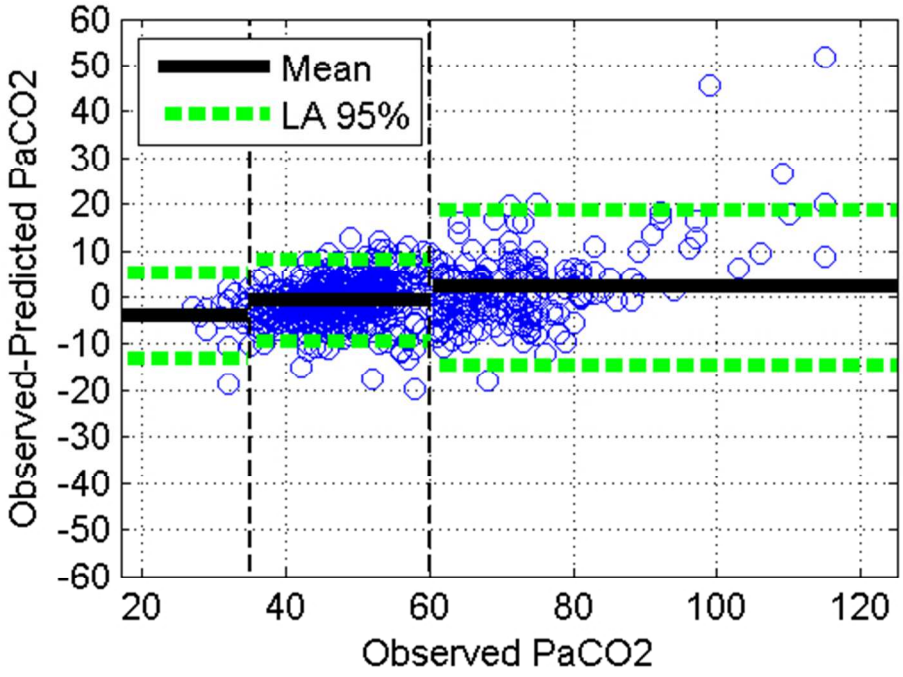

Figure 2b: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each PaCO2 bin for the MGP model. For PaCO2 <35 mean bias was -4.0 (95% Limits of Agreement (LA) +/-9.0); for PaCO2 between 35-60 mean bias was -0.8 (95%LA +/-8.7); for PaCO2 >60 mean bias was 2.1 (95% LA +/-16.6).
100x71mm (150 x 150 DPI)

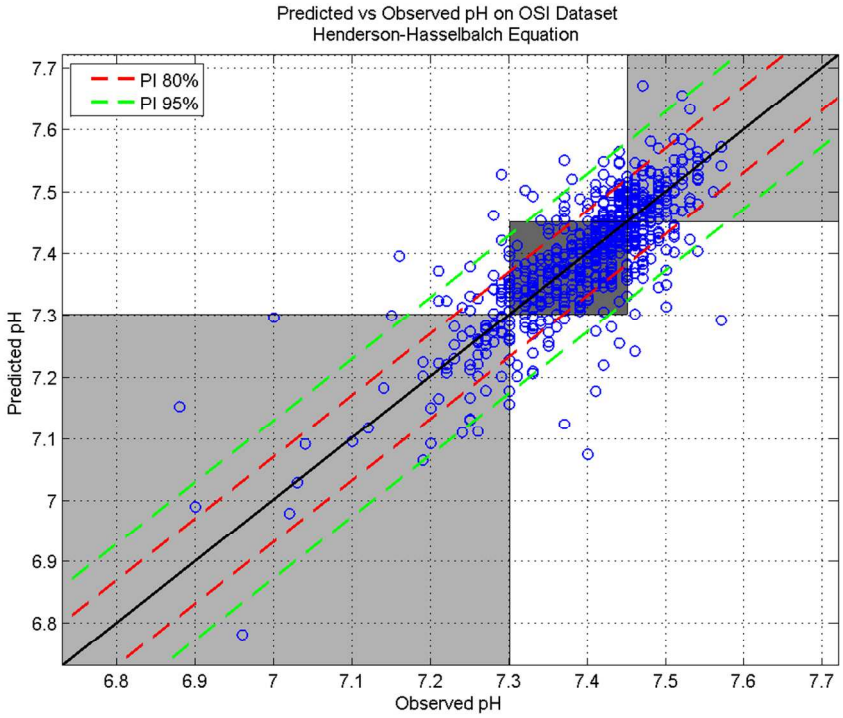

Figure 3a: Observed versus predicted values of pH based on the AVDSF equation to estimate PaCO2, and using the previous calculated bicarbonate value to plug into the Henderson-Hasselbalch equation. Red bounds are the 80% prediction intervals of +/- 0.07 and green bounds are the 95% prediction intervals of +/- 0.13. Shaded boxes represent the low, normal, and high bins. The points which lie in the shaded boxes would be classified into the correct bins (70% agreement, kappa = 0.48).
203x152mm (150 x 150 DPI)

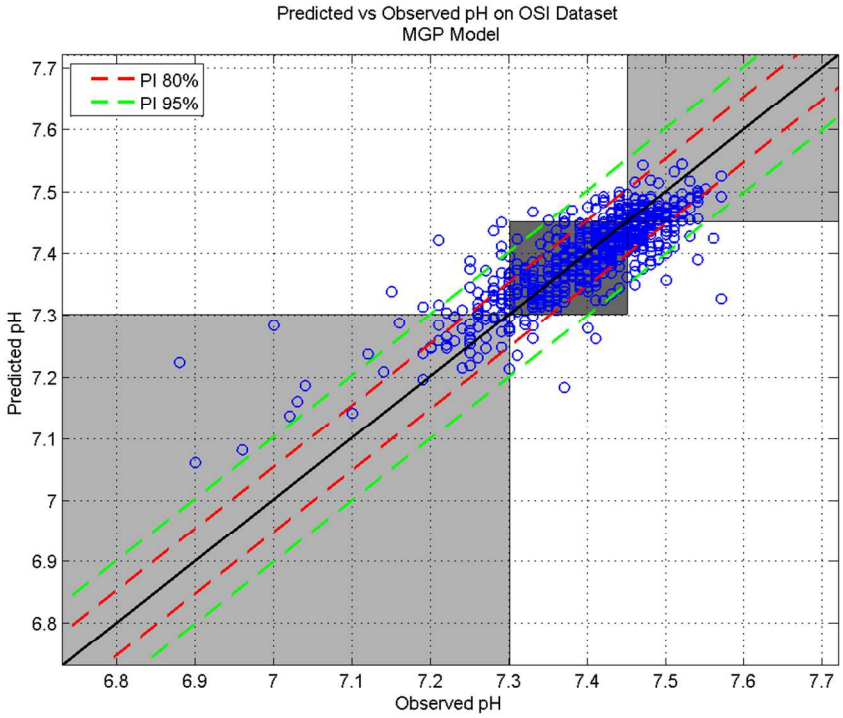

Figure 3b: Observed versus predicted values of pH based on the MGP model to estimate pH from the ETCO2, ventilator support, and previous known values of pH and PaCO2. Red bounds are the 80% prediction intervals of +/- 0.05 and green bounds are the 95% prediction intervals of +/- 0.10. Shaded boxes represent the low, normal, and high bins. The points which lie in the shaded boxes would be classified into the correct bins (72% agreement, kappa = 0.49).
203x152mm (150 x 150 DPI)

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

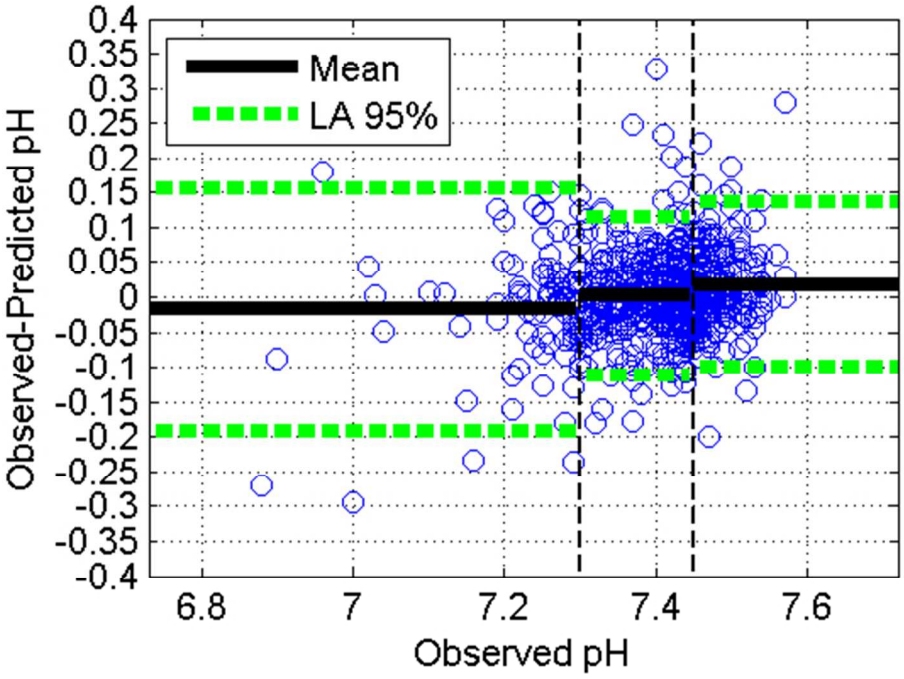

Figure 4a: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each pH bin based using the AVDSF equation to estimate PaCO2, and the previous calculated bicarbonate value to plug into the Henderson-Hasselbalch equation to estimate pH. For pH <7.3 mean bias was -0.017 (95% Limits of Agreement (LA) +/-0.17); for pH between 7.3-7.45 mean bias was 0.001 (95%LA +/-0.115); for pH≥ 7.45 mean bias was 0.015 (95% LA +/-0.113).
100x71mm (150 x 150 DPI)

RESPIRATORY CARE Paper in Press. Published on December 24, 2013 as DOI: 10.4187/respcare.02806

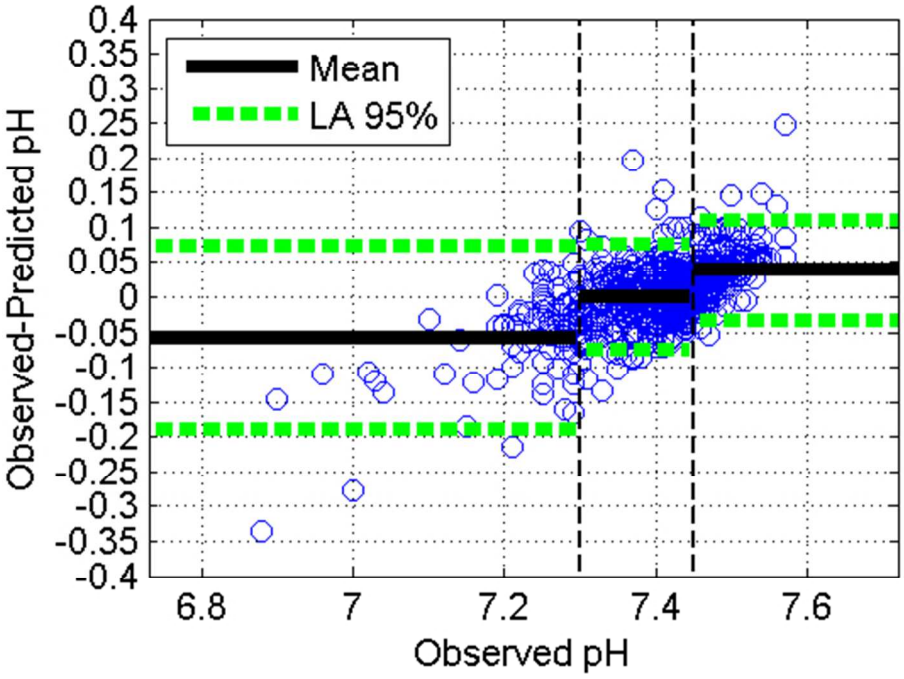

Figure 4b: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each pH bin for the MGP model. For pH <7.3 mean bias was -0.059 (95% Limits of Agreement (LA) +/-0.13); for pH between 7.3-7.45 mean bias was 0.001 (95%LA +/-0.078); for pH≥ 7.45 mean bias was 0.032 (95% LA +/-0.072).
100x71mm (150 x 150 DPI)