Algorithms to Estimate P_{aCO_2} and pH Using Noninvasive Parameters for Children with Hypoxemic Respiratory Failure

Robinder G Khemani MD MsCI, E Busra Celikkaya MSc, Christian R Shelton PhD, Dave Kale MSc, Patrick A Ross MD, Randall C Wetzel MB, and Christopher JL Newth MD

BACKGROUND: Ventilator management for children with hypoxemic respiratory failure may benefit from ventilator protocols, which rely on blood gases. Accurate noninvasive estimates for pH or P_{aCO_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 P_{aCO_2} from measures of ventilator support, pulse oximetry, and end-tidal carbon dioxide pressure (P_{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 data sets from previously published investigations. A baseline model estimated pH and P_{aCO_2} from P_{ETCO_2} using the previously observed relationship between P_{ETCO_2} and P_{aCO_2} or pH (using the Henderson-Hasselbalch equation). We developed a multivariate gaussian process (MGP) model incorporating other available noninvasive measurements. RESULTS: The training data set had 2,386 observations from 274 children, and the testing data set had 658 observations from 83 children. The baseline model predicted P_{aCO_2} within ± 7 mm Hg of the observed P_{aCO_2} 80% of the time. The MGP model improved this to ± 6 mm Hg. When the MGP model predicted P_{aCO_2} between 35 and 60 mm Hg, the 80% prediction interval narrowed to ± 5 mm Hg. The baseline model predicted pH within ± 0.07 of the 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 P_{aCO_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. *Key words: acute lung injury; pediatrics; respiration; artificial; capnography; decision support techniques.* [Respir Care 2014;59(8):1248–1257. © 2014 Daedalus Enterprises]

Introduction

Ventilator management for children with acute lung injury (ALI) varies widely.[1,2] Explicit ventilator protocols can standardize mechanical ventilation, provided that practitioners follow recommendations.[1,3-6] Specifically, potentially injurious ventilator settings are frequently not reduced, even with normal or over-ventilated pH or P_{aCO_2}.[1]

In general, in the acute phase of illness, ventilator settings are changed based on arterial blood gases (ABGs), requiring an arterial catheter and frequent blood samples, which is challenging in children.[7] Accurate and reliable noninvasive methods to estimate pH or P_{aCO_2} could allow for more frequent ventilator changes during the acute phase of illness to maintain permissive hypercapnia and to help in clinical decision making.

Pulse oximetry (S_{pO_2}) is routinely used for clinical decision making,[8] and clinicians change PEEP or F_{IO_2} in response to either P_{aO_2} or S_{pO_2} both in the acute phase of illness and during weaning. However, clinicians 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 P_{aCO_2}. The most widely used noninvasive sensor to estimate adequacy of ventilation is end-tidal carbon dioxide pressure (P_{ETCO_2}). However, the relationship between P_{ETCO_2} and P_{aCO_2} changes as a function of alveolar dead space. Additionally, estimating pH from P_{ETCO_2} is confounded by changing metabolic acidosis.

At the bedside, one can estimate P_{aCO_2} from P_{ETCO_2} using the alveolar dead-space fraction (AVDSF = $[P_{aCO_2} - P_{ETCO_2}]/P_{aCO_2}$).[9] Although 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 P_{ETCO_2} and P_{aCO_2}, to estimate future P_{aCO_2} from a known value of P_{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 P_{ETCO_2} for ventilator management have been applied to the weaning phase of mechanical ventilation.[14] We sought to develop a predictive algorithm to estimate pH and P_{aCO_2} that can account for changing alveolar dead space for application during the acute phase of illness.

We have previously demonstrated that noninvasive surrogates for intrapulmonary shunt (ie, oxygen saturation index [OSI] or S_{pO_2}/F_{IO_2}) are correlated with changes in AVDSF.[8,9] We hypothesize that incorporating noninvasive measures of intrapulmonary shunt, noninvasive continuously available values from the ventilator, and previously known values from the ABGs will permit development of a predictive algorithm to estimate P_{aCO_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.

Methods

We developed and tested algorithms using data sets from previous studies on children with acute hypoxemic respi-

QUICK LOOK

Current knowledge

Ventilator protocols based on measurements of arterial blood gases have proven useful in the management of pediatric hypoxemic respiratory failure. Accurate noninvasive measurements of arterial carbon dioxide could allow ventilator changes to optimize lung-protective ventilation without blood gas analysis.

What this paper contributes to our knowledge

In a group of mechanically ventilated pediatric patients with hypoxemic respiratory failure, a predictive model to estimate arterial carbon dioxide using the previously observed relationship between end-tidal carbon dioxide and P_{aCO_2} or pH predicted P_{aCO_2} within $\pm$ 6 mm Hg. The utility of these models in replacing arterial blood gases remains to be determined.

ratory failure.[1,8,9,15] We constructed data sets with simultaneous measurements of arterial pH, P_{aCO_2}, P_{aO_2}, pulse oximetry (when S_{pO_2} was $\le$ 97%), P_{ETCO_2} (measured via mainstream, with the same adapter size for all children as per ICU standards), and ventilator settings (mode, ventilator rate, peak inspiratory pressure, PEEP, exhaled tidal volume (mL/kg), and F_{IO_2}). We created composite variables for deficits in oxygenation, including the OSI (mean airway pressure $\times$ F_{IO_2} $\times$ $100/S_{pO_2}$) and S_{pO_2}/F_{IO_2}. We excluded measurements if there was a leak around the endotracheal tube of $\ge$ 20%,[16] if it had been $>$ 24 h since the previous ABG, or if there was only one ABG for an individual subject. The first ABG attained for each subject was used as baseline, and the algorithms generated estimates for pH and P_{aCO_2} at the time of subsequent ABGs. Predicted values for pH and P_{aCO_2} were compared with actual measured values. The study was approved by the Committee on Clinical Investigation at Children's Hospital Los Angeles with a waiver of informed consent (CCI-09-00126 and CCI-09-00287).

Data Set 1: Single-Center Data Set

We assembled this data set from a single-institution retrospective study. Children ($<$ 18 y of age) were included in this study if they were intubated and mechanically ventilated with at least one P_{aO_2}/F_{IO_2} $<$ 300 after intubation. Children with left ventricular dysfunction or cyanotic congenital heart disease were excluded. We have previously published the methods regarding data collection, subject characteristics, and ventilator support.[1,15] We extracted data from the electronic medical record, time-

ordered per subject. We extracted the closest charted value for S_{pO_2} and P_{ETCO_2}, which was at most 1 h before a study ABG. We have previously used this methodology to demonstrate that OSI, AVDSF, and S_{pO_2}/F_{IO_2} correlate with mortality.[9]

Data Set 2: Multi-Center Data Set

We assembled this data set from a 6-center prospective study in children. Children (<18 y of age) were included in this study if they were intubated and mechanically ventilated with an indwelling arterial line and $S_{pO_2} \leq 97\%$. Children with left ventricular dysfunction or cyanotic congenital heart disease were excluded. We have previously published the methods regarding data collection, subject characteristics, and ventilator support.[8] S_{pO_2} and P_{ETCO_2} were recorded prospectively precisely at the time of the ABG, with concurrent ventilator settings. Therefore, unlike with the single-center data set, P_{ETCO_2}, S_{pO_2}, ventilator settings, and ABG results were simultaneous. ABG values were not recorded if the pulse oximetry waveform was inadequate or if the subject had received endotracheal tube suctioning or invasive procedures for 30 min before the blood gas. P_{ETCO_2} was recorded when available as part of routine care. P_{ETCO_2} was not used routinely for all ventilated subjects in some of the study ICUs.

Analysis

We report the results of the statistical models trained on data set 1 and tested on data set 2. To predict both P_{aCO_2} and pH, we created 2 models. The first models used the previous simultaneous values for P_{aCO_2} and P_{ETCO_2} to calculate AVDSF, which was used to estimate the expected current value for P_{aCO_2}, based on a new P_{ETCO_2}. We used AVDSF instead of the difference between P_{aCO_2} and P_{ETCO_2} to control for proportionality inaccuracies as P_{aCO_2} increased. The first model for pH used this estimate for P_{aCO_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 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:

$$C(t,t') = \text{cov}[F(t),F(t')] \epsilon R^{nxn},$$

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 2 time points, and $S \epsilon R^{nxn}$ 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 $K \epsilon R^{nTxnT}$. We exploited the separable nature of our model and the simultaneity of the observations to avoid explicit computations with such a large matrix. We estimated the covariance matrix S from the training data (data set 1).

When testing on data set 2, we assumed we knew the measurements for all components except for the current values of P_{aCO_2} and pH. We predicted the mean and covariance of the marginal distributions of P_{aCO_2} and pH at the current time given all known measurements for all 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 P_{aCO_2} from the current minute ventilation and previous P_{aCO_2} and minute ventilation). It did not perform as well as AVDSF, so the results are not shown. We tested 2 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 P_{aCO_2}. To evaluate these outcomes, we generated 80% and 95% prediction intervals around the point estimate for pH and P_{aCO_2}. The purpose of this outcome was to assess whether the model could generate estimates that 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. P_{aCO_2} was binned: < 35, 35–60, and > 60 mm Hg. 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, and 80% and 95% prediction intervals for P_{aCO_2} or pH within each of the predicted bins. We also report the percentage of observa-

Table 1. Descriptive Statistics of Training and Testing Data Sets

	Training Data Set	Testing Data Set
Subjects, n	274	83
No. of observations	2,386	658
No. of observations per subject, median (IQR)	5 (2–11)	4 (2–8)
Age, mo, median (IQR)	17 (8–36)	36 (5–134)
Weight, kg, median (IQR)	4.6 (1.1–11.6)	13.5 (5.0–35.9)
Female, n (%)	115 (42)	27 (33)
Arterial blood gas, median (IQR)		
pH	7.37 (7.30–7.44)	7.42 (7.35–7.46)
P_{aCO_2}, mm Hg	49 (41–60)	51 (44–63)
P_{aO_2}, mm Hg	72 (61–88)	69 (61–81)
Time between ABGs, h	6.4 (3.6–11.7)	6.5 (2.0–12.4)
Noninvasive support		
S_{pO_2}	95 (92–97)	94 (92–96)
P_{ETCO_2}, mm Hg	38 (31–45)	40 (35–47)
Time between P_{ETCO_2} and ABG, min	28 (20–46)	Simultaneous
Ventilator settings		
Peak inspiratory pressure, cm H_2O	30 (25–35)	30 (26–37)
PEEP, cm H_2O	10 (6–12)	8 (5–10)
Mean airway pressure, cm H_2O	16 (12–20)	16 (13–20)
F_{IO_2}	0.60 (0.41–0.80)	0.55 (0.43–0.63)
Exhaled tidal volume, mL/kg	7.2 (5.8–8.8)	9 (7–11)
Ventilator frequency (breaths/min)	20 (16–26)	18 (16–24)
Lung disease severity		
Oxygen saturation index	10 (6–16)	9.5 (6.2–14)
Oxygenation index	12.6 (6.7–22.0)	12.6 (7.7–18.8)
S_{pO_2}/F_{IO_2}	162 (119–218)	166 (145–216)
P_{aO_2}/F_{IO_2}	127 (86–192)	130 (95–178)
AVDSF ($[P_{aCO_2} - P_{ETCO_2}]/P_{aCO_2}$)	0.24 (0.14–0.34)	0.22 (0.15–0.30)

IQR = interquartile
ABG = arterial blood gas
P_{ETCO_2} = end-tidal carbon dioxide pressure
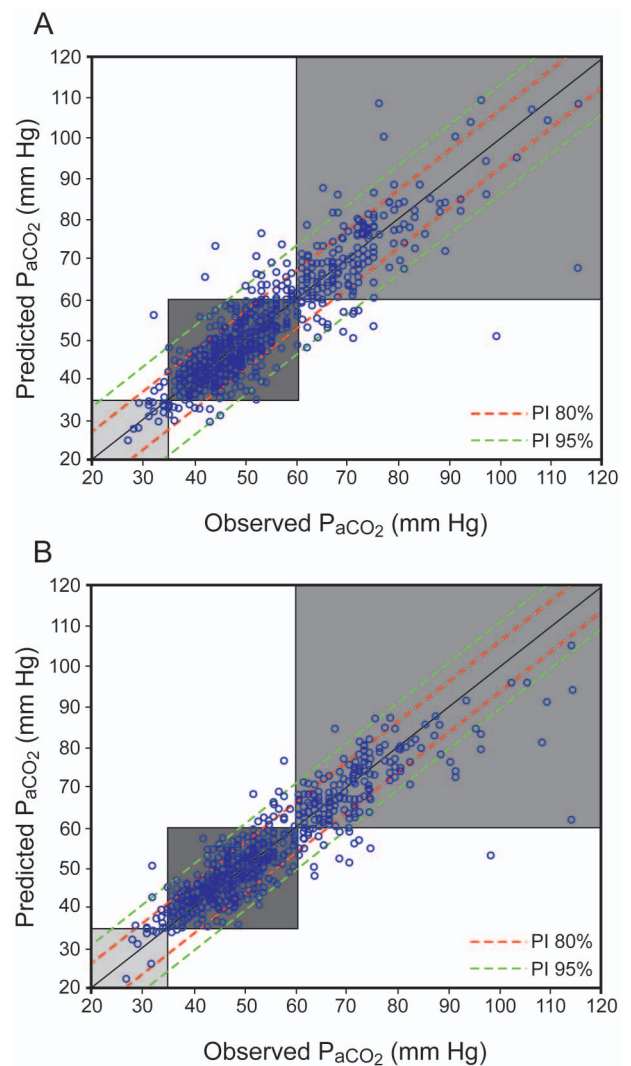AVDSF = alveolar dead-space fraction

Fig. 1. A: Observed versus predicted values of P_{aCO_2} based on the alveolar dead-space fraction equation to estimate P_{aCO_2} from the current value of P_{ETCO_2}, knowing the previous relationship between P_{aCO_2} and P_{ETCO_2}. Red bounds are the 80% prediction intervals (PI) of $\pm$ 7 mm Hg, and green bounds are the 95% prediction intervals of $\pm$ 13 mm Hg. Shaded boxes represent the low, normal, and high bins. The points that lie in the shaded boxes would be classified into the correct bin (89% agreement, kappa = 0.76). B: Observed versus predicted values of P_{aCO_2} based on the multivariate gaussian process model to estimate P_{aCO_2} from P_{ETCO_2}, ventilator support, and previous known values of pH and P_{aCO_2}. Red bounds are the 80% prediction intervals of $\pm$ 6 mm Hg, and green bounds are the 95% prediction intervals of $\pm$ 11 mm Hg. Shaded boxes represent the low, normal, and high bins. The points that lie in the shaded boxes would be classified into the correct bins (91% agreement, kappa = 0.80).

tions that fall within Clinical Laboratory Improvement Amendments (CLIA) standards for P_{aCO_2} (the greater of $\pm$ 5 mm Hg or 8%) and pH ($\pm$ 0.04).[19]

Results

Of the 398 children enrolled in the single-center retrospective study (data set 1), 274 met inclusion criteria, with $S_{pO_2} \leq$ 97% and P_{ETCO_2} results available at most 1 h before the ABG. Of the 103 children without cyanotic congenital heart disease enrolled in the multi-center study (data set 2), 83 met inclusion criteria, with P_{ETCO_2} data available at the time of the ABG. Hence, this training data set (data set 1) had 2,386 observations (aligned P_{ETCO_2}, S_{pO_2}, ABG, and ventilator data) from 274 children. The testing data set (data set 2) had 658 observations from 83

children. In general, the data sets were similar with respect to disease severity, blood gas parameters, and ventilator support. The subjects had moderate-to-severe lung injury, with a median F_{IO_2} of 0.6 (interquartile range of 0.4–0.8), a median P_{aO_2}/F_{IO_2} of 127 (interquartile range of 86–192),

and a median oxygenation index of 12.6 (interquartile range of 6.7–22) in data set 1. Results for data set 2 were similar (Table 1). The median time between observations (ABGs) was 6.5 h in both data sets (Table 1).

Model Training (Data Set 1)

The model was trained (parameters were estimated) on data set 1. Variables included in the model were P_{ETCO_2}, OSI, peak inspiratory pressure, PEEP, ventilator rate, tidal volume, minute ventilation, an interaction term of P_{ETCO_2}/min of ventilation, dynamic compliance of the respiratory system, and previous values for pH, P_{aCO_2}, and P_{ETCO_2}. Other variables considered but not used in the model were pressure support and S_{pO_2}/F_{IO_2}. This is because pressure support varied little (the value was almost always 10 cm H_2O) and S_{pO_2}/F_{IO_2} was included in the calculation of OSI, which was included in the model.

Prediction of P_{aCO_2} in the Testing Data Set (Data Set 2)

We constructed a baseline model to predict P_{aCO_2} using the previous AVDSF with the current P_{ETCO_2}. The predicted values were on average 0.3 ± 7.2 mm Hg (mean $\pm$ SD) higher than the observed values. Overall, 67.5% of the predicted P_{aCO_2} would fall within CLIA standards (greater of $\pm$ 5 mm Hg or 8% difference) against the measured P_{aCO_2}. Eighty percent of the predicted values were within $\pm$ 7 mm Hg of the observed values, and 95% were within $\pm$ 13 mm Hg (Fig. 1A). When binning the observed and predicted P_{aCO_2}, the overall agreement was 89%, with a kappa of 0.76 (Table 2). The accuracy was best in the normal or low P_{aCO_2} bins, where 80% of the predicted values were within $\pm$ 6 mm Hg of the observed values, and 95% were within $\pm$ 12 mm Hg (Table 3 and Fig. 2A).

The MGP model derived from data set 1 performed slightly better than the baseline model using AVDSF. The predicted values were on average 0.02 ± 6.1 mm Hg (mean $\pm$ SD) higher than the observed values. Overall, 73.6% of the predicted P_{aCO_2} would fall within CLIA standards against the measured P_{aCO_2}. For the MGP model, 80% of the predicted values were within $\pm$ 6 mm Hg of the observed values and 95% were within $\pm$ 11 mm Hg (see Fig. 1B). When binning the observed and predicted P_{aCO_2}, the overall agreement was 91%, with a kappa of 0.80 (see Table 2). Within each P_{aCO_2} bin, the prediction intervals were narrower than with the AVDSF model, and in the normal P_{aCO_2} bin, 80% of the predicted values were within $\pm$ 5 mm Hg of the observed values, and 95% were within $\pm$ 10 mm Hg. In the low P_{aCO_2} bin, 80% of the

Table 2. Observed Versus Predicted Bins of P_{aCO_2} Generated From the AVDSF and MGP Models for P_{aCO_2}

Predicted P_{aCO_2}	Actual P_{aCO_2}			Total
	< 35 mm Hg	35–60 mm Hg	> 60 mm Hg	
AVDSF model				
< 35 mm Hg	13 (52)	12 (48)	0	25 (4)
35–60 mm Hg	5 (1)	402 (93)	26 (6)	433 (66)
> 60 mm Hg	0	30 (15)	170 (85)	200 (30)
Total	18	444	196	658
MGP model				
< 35 mm Hg	8 (73)	3 (27)	0	11 (2)
35–60 mm Hg	10 (2)	426 (92)	30 (6)	466 (71)
> 60 mm Hg	0	15 (8)	166 (92)	181 (30)
Total	18	444	196	658

The overall agreement for the alveolar dead-space fraction (AVDSF) model was 89% with a kappa of 0.76 compared with 91% with a kappa of 0.80 for the multivariate gaussian process (MGP) model. The values in parentheses are percentages calculated across the rows to represent the percentage of predicted values in which the actual P_{aCO_2} fell in each bin. For example, for the AVDSF model, P_{aCO_2} was predicted to be between 35 and 60 mm Hg 433 times, and 402 (93%) of these times, the actual P_{aCO_2} was also between 35 and 60 mm Hg. The MGP model has better agreement between observed and predicted P_{aCO_2} bins than the AVDSF model, particularly when P_{aCO_2} is estimated to be < 35 or > 60 mm Hg.

Table 3. The 80% and 95% PIs for the AVDSF and MGP Models to Predict P_{aCO_2} and the AVDSF Model Using the Henderson-Hasselbalch Equation and the MGP Model to Predict pH

	AVDSF Model		MGP Model	
	80% PI	95% PI	80% PI	95% PI
Predicted P_{aCO_2}				
< 35 mm Hg	6	12	3	4
35–60 mm Hg	6	12	5	10
> 60 mm Hg	9	17	8	15
	AVDSF/Henderson-Hasselbalch Model		MGP Model	
	80% PI	95% PI	80% PI	95% PI
Predicted pH				
< 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

For the alveolar dead-space fraction (AVDSF) model, when P_{aCO_2} is between 35 and 60 mm Hg, 80% of predicted values for P_{aCO_2} would fall within $\pm$ 6 mm Hg of the observed values. For the multivariate gaussian process (MGP) model, the prediction intervals are narrower than the AVDSF model in all 3 ranges of predicted P_{aCO_2}. For the AVDSF model using the Henderson-Hasselbalch equation, when pH is between 7.30 and 7.44, 80% of predicted values for pH would fall within $\pm$ 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.
PI = prediction interval

predicted values were within $\pm$ 3 mm Hg of the observed values, and 95% were within $\pm$ 4 mm Hg (see Table 3 and Fig. 2B).
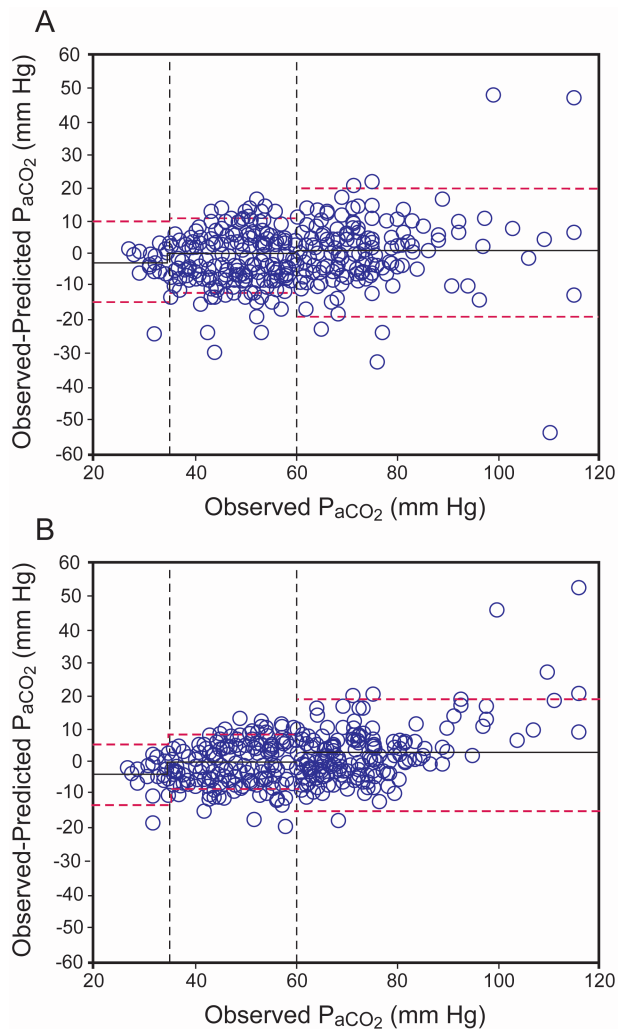
Fig. 2. A: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each P$_{aCO_2}$ bin for the alveolar dead-space fraction model. For P$_{aCO_2}$ < 35 mm Hg, the mean bias was − 2.69 (95% limits of agreement, ± 11.8); for P$_{aCO_2}$ = 35–60 mm Hg, the mean bias was −0.5 (95% limits of agreement, ± 11.1); and for P$_{aCO_2}$ > 60 mm Hg, the mean bias was 0.6 (95% limits of agreement, ± 19.5). B: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each P$_{aCO_2}$ bin for the multivariate gaussian process model. For P$_{aCO_2}$ < 35 mm Hg, the mean bias was −4.0 (95% limits of agreement, ± 9.0); for P$_{aCO_2}$ = 35–60 mm Hg, the mean bias was −0.8 (95% limits of agreement, ± 8.7); and for P$_{aCO_2}$ > 60 mm Hg, the mean bias was 2.1 (95% limits of agreement, ± 16.6). The vertical dashed lines denote each P$_{aCO_2}$ bin, the thick black line shows the mean, and the red dashed lines denote 95% limits of agreement.
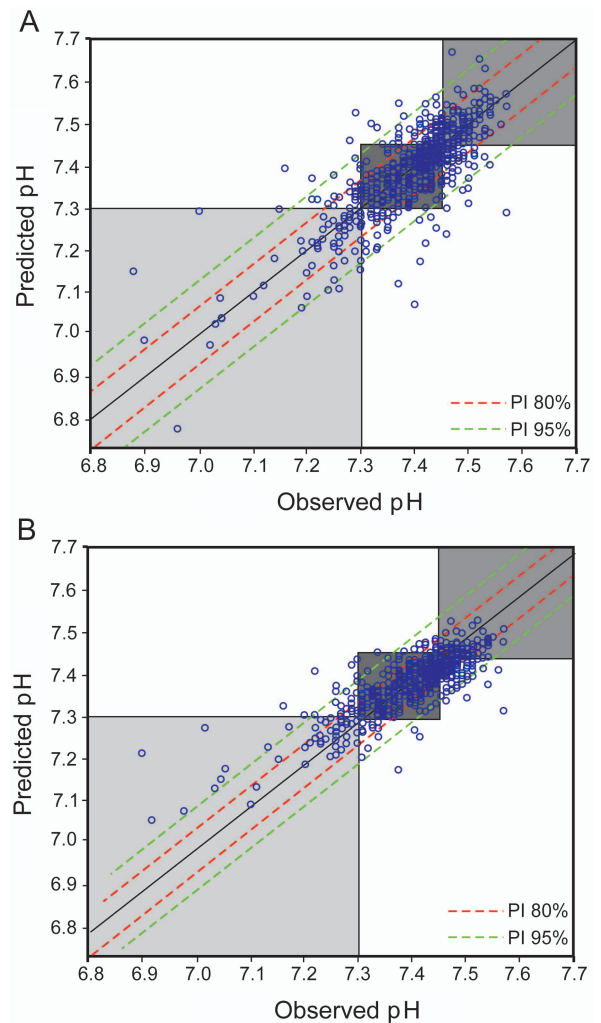
Fig. 3. A: Observed versus predicted values of pH based on the alveolar dead-space fraction equation to estimate P$_{aCO_2}$ and using the previous calculated bicarbonate value to plug into the Henderson-Hasselbalch equation. Red bounds are the 80% prediction intervals (PI) of ± 0.07, and green bounds are the 95% prediction intervals of ± 0.13. Shaded boxes represent the low, normal, and high bins. The points that lie in the shaded boxes would be classified into the correct bins (70% agreement, kappa = 0.48). B: Observed versus predicted values of pH based on the multivariate gaussian process model to estimate pH from P$_{ETCO_2}$, ventilator support, and previous known values of pH and P$_{aCO_2}$. 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 that lie in the shaded boxes would be classified into the correct bins (72% agreement, kappa = 0.49).

Prediction of pH in the Testing Data Set (Data Set 2)

We used the predicted P$_{aCO_2}$ from the AVDSF model with the calculated serum bicarbonate from the previous ABG with the Henderson-Hasselbalch equation to predict pH. The predicted values were on average 0.004 ± 0.064 (mean ± SD) lower than the observed values. Overall,

59.3% of the 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 the observed values, and 95% were within ± 0.13 (Fig. 3A). When binning the observed and predicted pH, the overall agreement between the observed and predicted bins was 70%, with a kappa of 0.48 (Table 4). The best accuracy

Table 4.   Observed Versus Predicted Bins of pH Generated From the AVDSF Model Using the Henderson-Hasselbalch Equation and the MGP Model for pH

Predicted pH	Actual pH			Total
	< 7.3	7.3–7.44	≥ 7.45	
AVDSF model using the Henderson-Hasselbalch equation				
< 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
MGP model				
< 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

The overall agreement for the alveolar dead-space fraction (AVDSF) model using the Henderson-Hasselbalch equation was 70% with a kappa of 0.48 compared with 72% with a kappa of 0.49 for the MGP model. The values in parentheses are percentages 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 the Henderson-Hasselbalch equation, the 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.
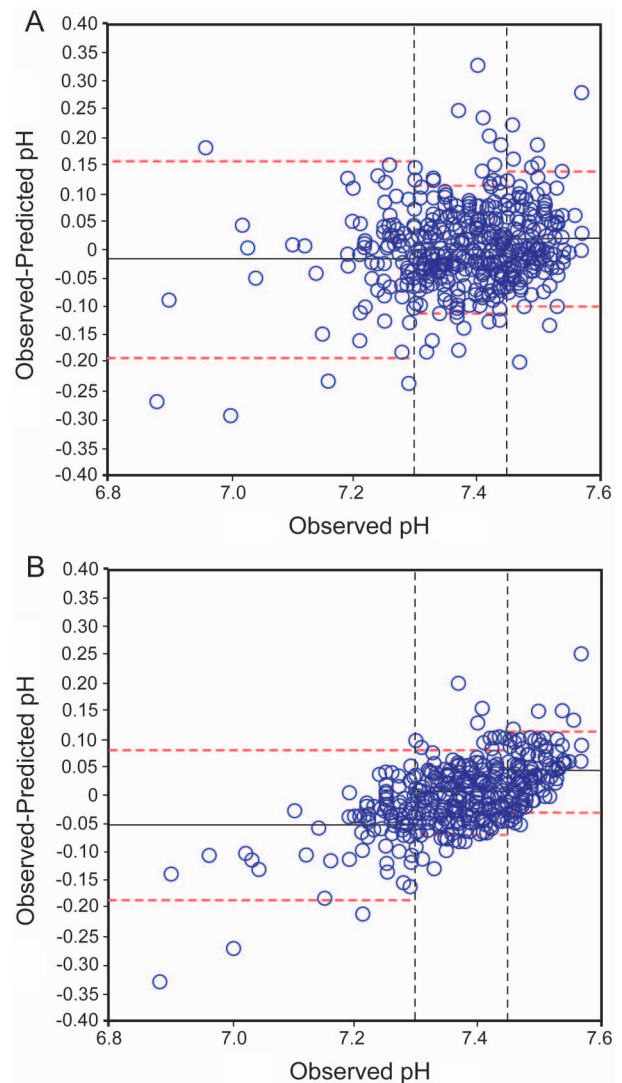

Fig. 4. A: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each pH bin based using the alveolar dead-space function equation to estimate P$_{aCO_2}$ and the previous calculated bicarbonate value to plug into the Henderson-Hasselbalch equation to estimate pH. For pH < 7.3, the mean bias was −0.017 (95% limits of agreement, ± 0.17); for pH between 7.3 and 7.45, the mean bias was 0.001 (95% limits of agreement, ± 0.115); and for pH ≥ 7.45, the mean bias was 0.015 (95% limits of agreement, ± 0.113). B: Bland-Altman plot demonstrating mean bias and 95% limits of agreement as a function of each pH bin for the multivariate gaussian process model. For pH < 7.3, the mean bias was −0.059 (95% limits of agreement, ± 0.13); for pH between 7.3 and 7.45, the mean bias was 0.001 (95% limits of agreement, ± 0.078); and for pH ≥ 7.45, the mean bias was 0.032 (95% limits of agreement, ± 0.072). The vertical dashed lines denote each pH bin, the thick black line shows the mean, and the red dashed lines denote 95% limits of agreement.

was in the normal pH bin, where 80% of the predicted values were within ± 0.06 of the observed values, and 95% were within ± 0.10 (see Table 3 and Fig. 4A).

The same MGP model (trained on data set 1) was used to predict pH for data set 2. In general, the MGP model was slightly superior to the model using AVDSF and the Henderson-Hasselbalch equation. The predicted values were on average 0.002 ± 0.05 (mean ± SD) 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 the observed values, and 95% were within ± 0.10 (see Fig. 3B). When binning the observed and predicted pH, the overall agreement between the observed and predicted bins was 72%, with a kappa of 0.49 (see Table 4). Within each pH bin, the prediction intervals were narrower than with the Henderson-Hasselbalch model, and in the normal pH bin, 80% of the predicted values were within ± 0.05 of the observed values, and 95% were within ± 0.10. In the high pH bin, 80% of the predicted values were within ± 0.05 of the observed values, and 95% were within ± 0.07 (see Table 3 and Fig. 4B).

Alternative Training and Testing Data Sets

We repeated the analysis using data set 2 as the training data set and data set 1 as the testing data set. In general,

each model had larger prediction intervals and less agreement between observed and predicted bins of pH and P$_{aCO_2}$ (data not shown).

Discussion

We have demonstrated a first step in the application of machine-learning algorithms to estimate pH and P_{aCO_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, although models that use previous known relationships between P_{ETCO_2} and P_{aCO_2} perform reasonably well, we can modestly improve the accuracy by incorporating noninvasive markers of oxygenation as well as changes to ventilator settings. It may be that these are surrogates for changes in alveolar dead space. Our MGP model predicts P_{aCO_2} with 80% prediction intervals of ± 6 mm Hg and pH with 80% prediction intervals of ± 0.05. It performed best in the middle or over-ventilated range of pH and P_{aCO_2}. For example, if the model predicted P_{aCO_2} to be 50 mm Hg, 80% of the time, the actual P_{aCO_2} (if one were to draw an ABG) would be between 45 and 55 mm Hg, and 95% of the time, it would be between 40 and 60 mm Hg. 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 P_{aCO_2} with CLIA-acceptable equivalence to a blood gas machine 74% of the time and pH 67% of the time.

Some may believe that the confidence in these predictions is not adequate for clinical decision making. Although 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 $\pm 10\%$ against co-oximetry.[20] For example, if the pulse oximeter reading is 85%, then 50% of the time, the actual S_{aO_2} on co-oximetry would lie between 75 and 83%, but for 95% certainty, the S_{aO_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. Although S_{pO_2} is more accurate in the range frequently seen for children with ALI, this example is meant to illustrate that the parameters practitioners routinely use for clinical decisions relating to mechanical ventilation may have more uncertainty than 20%. As such, 80% certainty that the P_{aCO_2} is in a range of 10 mm Hg or that the pH is in a range of 0.1 (as seen in our model in the normal or over-venti-

lated 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 that there are several potential applications of these algorithms. First, noninvasive estimates of P_{aCO_2} or pH may decrease the number of ABGs. If clinicians can be 80% confident that the P_{aCO_2} lies within a range of 10 mm Hg or that the pH lies within a range of 0.1, they 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 estimated open loop allows providers 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 that requires an assessment every 2–4 h. The decision support tool could display an estimate for the predicted pH, with a prediction interval. Clinicians could accept or reject the protocol's recommendation or obtain an ABG if they are 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 P_{aCO_2} bins). We have previously demonstrated there are many lost opportunities to be lung-protective when clinicians do not wean the ventilator even when settings are high and the pH is normal or high.[1] A continuously available estimate of pH or P_{aCO_2} may make clinicians more willing to wean ventilator settings 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 that assume no change in dead space (such as the AVDSF model) and those that try to capture surrogates for changing dead space (ie, the MGP model) would perform even better in children with minimal or no lung injury. This should be tested. The median AVDSF in these 2 data sets was ~ 0.23, in line with our previous publication on AVDSF demonstrating that such values are independently associated with mortality.[9] As such, these data sets represent the

lung-injured children we often take care of in our ICUs, for whom lung-protective ventilation has the potential to improve outcomes.

Although 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 P_{aCO_2}. However, machine-learning techniques have been used extensively in medicine,[21] in gene expression studies,[22-24] for classification of cardiac arrhythmias,[25] for predicting morbidity after coronary artery bypass surgery,[26] and for 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 although 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 data set and tested on the single-center data set. We believe that this is a function of the simultaneous assessments of S_{pO_2}, P_{ETCO_2}, and ABG values in the prospective multi-center data set compared with the retrospective single-center data set. Because the real-time application of the algorithm will be with controlled, continuously available data, we believe that it will more likely perform as it did with the multi-center data. This needs to be tested. Third, we elected not to split the multi-center data set into training and testing data because we were worried about sample size. Fourth, the testing data set was relatively small (84 subjects), and the algorithm should be evaluated in another group of subjects. Fifth, there were 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 the AVDSF model and may perform differently when applied to a different validation data set. 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.

We may be able to improve model accuracy with temporally continuous values for P_{ETCO_2} and S_{pO_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

Noninvasive, continuously available measurements and ventilator settings may be helpful for predictive algorithms that estimate P_{aCO_2} and pH for children with hypoxemic respiratory failure. The current level of accuracy offers some applications, particularly for standardizing decisions about decreasing ventilator support in line with lung-protective strategies when the pH or P_{aCO_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.

REFERENCES

1. Khemani RG, Sward K, Morris A, Dean JM, Newth CJ, NICHD Collaborative Pediatric Critical 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 CL, 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. Ventilatory management of acute lung injury and acute respiratory distress syndrome. JAMA 2005; 294(22):2889-2896.
4. The Acute Respiratory Distress Syndrome Network. 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 CJ. 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 MA. 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. Crit Care Med 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. Crit Care Med 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. 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, Cheifetz IM. 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. 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 3rd, 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 C, Koller D. Continuous time Bayesian networks. In: Darwiche, A, Friedman, N. Proceedings of the Eighteenth International Conference on Uncertainty in Artificial Intelligence, August 1-4, 2002, Alberta, Canada. San Francisco: Morgan Kaufmann; 2002;378-387.

18. Khemani RG, Kale D, Ross PA, Wetzel RC, Newth C. Predicting serum pH using non-invasive measurements to enable decision support for mechanically ventilated children with acute lung injury (abstract). http://www.atsjournals.org/doi/abs/10.1164/ajrccm-conference.2012.185.1_MeetingAbstracts.A2358. *Accessed June 23, 2014.*

19. Clinical Laboratory Improvement Amendments. CLIA proficiency testing criteria for acceptable analytic performance. Fed Regist 1992; 57(40):7002-7186.

20. Ross PA, Newth CJ, Khemani RG. Accuracy of pulse oximetry in children. Pediatrics 2014;133(1):22-29

21. Meyfroidt G, Güiza F, Ramon J, Bruynooghe M. 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. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. Bioinformatics 2005; 21(5):631-643.

23. Zhao X, Cheung LW. 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. 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. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. Artif Intell Med 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. Crit Care 2006;10(3):R94.

27. Guiza Grandas, F, Blockeel H, Bruynooghe M, Van Loon K, Aerts JM, Berckmans D, et al. Time-series analysis techniques combined with Gaussian process classifiers for prediction of clinical stability after coronary bypass surgery. In: Proceedings of the Sixth IASTED International Conference on Biomedical Engineering, Innsbruck, Austria, February 13-15 2008. Anaheim, CA: ACTA Press; 2008;216-221.

28. Ganzert S, Kramer S, Möller K, Steinmann D, Guttmann J. Prediction of mechanical lung parameters using Gaussian process models. Artif Intell Med 2009;5651:380-384.