Continuous Noninvasive Blood Gas Estimation in Critically Ill Pediatric Patients With Respiratory Failure

Junzi Dong (junzi.dong@philips.com)
Philips Research Cambridge: Philips Research https://orcid.org/0000-0002-0382-1141
Minnan Xu-Wilson
Philips Research Cambridge: Philips Research
Bryan R. Conroy
Philips Research Cambridge: Philips Research
Robinder G. Khemani
University of Southern California Keck School of Medicine
Christopher J.L. Newth
University of Southern California Keck School of Medicine

Research

Keywords: blood gas monitoring, decision support techniques, machine learning, ventilation

Posted Date: November 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-111861/v1

Continuous Noninvasive Blood Gas Estimation in Critically Ill Pediatric Patients with Respiratory Failure

Junzi Dong PhD[1], Minnan Xu-Wilson PhD[1], Bryan R. Conroy PhD[1], Robinder G. Khemani MD[2,3], Christopher J.L. Newth MD, FRCPC[2,3]

Affiliations: [1]Connect Care and Personal Health, Philips Research North America, Cambridge, MA; [2]Department of Anesthesiology and Critical Care Medicine, Children's Hospital Los Angeles, Los Angeles, CA; [3]Department of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, CA.

Corresponding author:
Junzi Dong
junzi.dong@philips.com
222 Jacobs Street
Cambridge, MA 02141
314 452 0075

Background

Patients supported by mechanical ventilation require frequent invasive blood gas samples to monitor and adjust the level of support. We developed a transparent and novel blood gas estimation model to provide continuous monitoring of blood pH and arterial CO_2 in between gaps of blood draws, using only readily available noninvasive data sources in ventilated patients.

Methods

The model was trained on a derivation dataset (1,883 patients, 12,344 samples) from a tertiary pediatric intensive care center, and tested on a validation dataset (286 patients, 4,030 samples) from the same center obtained at a later time. The model uses pairwise non-linear interactions between predictors and provides point-estimates of blood gas pH and arterial CO_2 along with a range of prediction uncertainty.

Results

The model predicted within Clinical Laboratory Improvement Amendments of 1988 (CLIA) acceptable blood gas machine equivalent in 74% of pH samples and 80% of PCO_2 samples. Prediction uncertainty from the model improved estimation accuracy by identifying and abstaining on a minority of high-uncertainty samples.

Conclusions

The proposed model estimates blood gas pH and CO_2 accurately in a large percentage of samples. The model's abstention recommendation coupled with ranked display of top predictors for each estimation lends itself to real-time monitoring of between gaps of blood draws, and the model may be used to help users determine when a new blood draw is required and delay blood draws when not needed.

Key words: blood gas monitoring, decision support techniques, machine learning, ventilation

BACKGROUND

Patients in severe respiratory distress are often supported by intubation with mechanical ventilation. The correct level of ventilation is critical for life support without further lung injury. Blood gas pH and arterial CO_2 pressure ($PaCO_2$) obtained through invasive blood draws are relied upon to help determine ventilator settings. In the acute phase of injury, frequent blood draws are needed to determine blood gases[1]. This is especially difficult in pediatric patients where arterial access, pain, and blood loss are major concerns[2]; moreover, arterial catheters are an under-recognized source of infection[3]. Improvements in pulse oximetry providing continuous monitoring of oxygenation has proved helpful in children and shifted practice patterns in pediatric intensive care to reduce use of arterial catheters[1,2]. With respect to ventilation, exhaled CO_2 monitored through capnography is correlated with blood gas (BG) CO_2 tension but has not been able to provide the type of continuous monitoring oximetry does. However, the frequency of BG sampling is decreased with capnography usage[4-6], demonstrating that clinicians informally use capnography to determine the direction of blood pH changes.

There has long been interest in estimating BG pH and PCO_2 from end-tidal CO_2 ($PetCO_2$)[7-9] and over the past few years there have been some stimulating new investigations on estimating these in pediatric patients noninvasively[10-13]. These studies show that $PetCO_2$ concentrations along with other noninvasive measurements can be used to estimate the values of blood pH and PCO_2 without taking an invasive blood sample. Nonetheless, challenges to clinical adoption remain. Prediction accuracy outside the normal pH range is low[10,11], and there is a lack of clinical confidence in the predicted values.

The goal of this study is to develop continuous BG estimation that is accurate in all pH ranges for mechanically ventilated children with a wide range of severity of lung injury and hemodynamic

support. Special consideration was given to develop a model suitable for clinical adoption. Estimations are made with a prediction uncertainty range, and the model can abstain from making inaccurate estimations using the prediction uncertainty in case of large physiological fluctuations. Investigations on estimation accuracy over time provides guidance on the timeframe in which continuous noninvasive monitoring can be used. To further help users interpret and understand an estimated BG value, predictors are ranked by those with most significant contributions to the estimated value and displayed. We developed the model on a large derivation dataset spanning five years of data using novel modeling techniques and tested it on unseen validation data.

METHODS

Following the recommendations of the Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis[14], we developed and validated a BG estimation model that either provides an estimate of the current pH and PCO_2 or abstains from estimation.

Study population

The derivation and validation datasets were collected from pediatric critical care patients admitted to a tertiary pediatric intensive care center. Figure 1 illustrates data extraction steps for both cohorts. Derivation and validation cohorts spanned different times, and samples from the same patient could not appear in both cohorts.

Derivation cohort extraction

The derivation dataset was collected from patient measurements made between September 2012 and May 2017 and stored prospectively in the hospital's dedicated critical care SQL Server (Microsoft, Redmond, WA). A dataset containing BG, granular physiological and ventilator data collected within ±1 minute of BG sample time was extracted from these medical records. pH and PCO_2 measurements were obtained from both arterial and capillary blood gases, which made up 90% and 10% of the data samples, respectively. In model development and analyses, arterial and capillary BG were used inter-changeably given closeness of capillary BG to arterial BG[15]. Samples with missing information in $PetCO_2$ measurement or medical record number (MRN) were removed. Derivation data was resampled to balance pH distribution and improve model performance in sparsely represented pH regions (eFig. 1). Patients on extracorporeal membrane oxygenation support were removed. A plausibility check was performed on measurement values as shown in eTable 1. Samples for the same patient were linked in time, and samples without a prior BG within 24 hours were removed. The final derivation cohort was split into 5 outer- and 5 inner- folds for cross-validation (CV) using nested CV[16] stratified by pH. Predictor and model selection were performed on inner CV folds, and final training was done on all folds.

Validation cohort extraction

The validation dataset was collected from measurements made between June to December 2017. The pH resampling step was not done for the validation dataset, in order for validation performance to reflect a natural BG distribution. Patients in the validation cohort who had already appeared in the derivation cohort were removed.

Predictors and target

Target variables $pH[t], PaCO_2[t]$
Predictor variables $pH[t-1]$, $PaCO_2[t-1]$, $HCO_3^-[t-1]$, $etCO_2[t]$, $etCO_2[t-1]$, ΔFiO_2, $\Delta PEEP$, ΔPIP, $\Delta MnAwP$, ΔSpO_2, $\Delta TVin$, $\Delta TVexp$

Table 1 *Targets for prediction, and predictors included in the final model.*

Suppose the current time is t and the previous BG was measured at time $t-1$. The target and predictor variables are shown in Table 1. 'Delta' predictors are the difference between current (taken at time t) and previous ($t-1$) measurements, e.g., $\Delta SpO_2 = SpO_2\,[t] - SpO_2[t-1]$. Final predictors included in the model were selected from 22 candidate predictors (eTable 2). Predictors were selected on inner CV folds using ridge regression to remove co-linearity between predictors and spurious correlations between predictors and targets.

Statistical analysis

Model building

A novel pairwise regression model was developed to model interactions between one key predictor (previous pH) and non-key predictors. This model allows differences in physiology between patients in different pH ranges to be modeled independently while representing monotonic relationships between non-key predictors and BG. The model is mathematically expressed as

$$\hat{y} = \sum_{j=1}^{M} \sum_{i=1}^{K} w_{i,\,j} \cdot f_j(z) \cdot x_i,$$

where $\hat{y}$ denotes the predicted target, $x_i, i = 1, \ldots, K$ denotes K non-key predictors, z denotes the key predictor, and $f_j(z),\ j = 1, \ldots, M$ denotes sigmoid functions centered at M different values of the key predictor.

The model is interpretable: pairwise interactions can be easily visualized[17] as show in Fig. 2 and the contribution of each predictor can be separated to generate predictor importance rankings for each estimation. Data processing, modeling, and analyses were performed in Python.

Prediction uncertainty around point-estimates

Prediction uncertainty was modeled using bootstrap estimation of uncertainty[18], by building separate models for each derivation CV fold and quantifying the agreement between them. For a given sample, the prediction uncertainty is the variance in predicted target values by all separate models. An acceptable threshold for prediction uncertainty was determined after training, so that only samples with uncertainty lower than the threshold generated an estimation. Estimations were abstained on samples with high uncertainty.

The point-estimate of BG is generated from a final model trained on all derivation data. Separate models from derivation CV folds provide a prediction uncertainty range around the point-estimate.

Predictor importance ranking

When an estimate is made, predictor contribution to the estimation is ranked. The model can be rewritten as $\hat{y} = \sum_{i=0}^{K} g(x_i)$, which denotes the sum of contributions from individual predictors $g(x_i) = w_{i,j} \cdot f_j(z) \cdot x_i$.

Given a sample $\vec{x} = [x_0, x_1, \ldots, x_i, \ldots, x_K]$, the importance of the ith predictor is

$$I_i = f_i(x_i) - f_i(\bar{x}_i),$$

where $\bar{x}_i$ denotes the population mean of the predictor. When x_i is close to $\bar{x}_i$, I_i will be equal or close to 0, which means predictor x_i contributes little to the overall estimate. When x_i deviates from the population mean, I_i shifts away from 0 to highlight the increased contribution of x_i.

Baseline models

The alveolar dead-space fraction (AVDSf) model was used to establish a baseline for comparison. It uses the alveolar dead-space fraction ($AVDSf = (PaCO_2 - PetCO_2)/PaCO_2$) calculated from the previous BG to estimate the current PCO_2. This estimate for PCO_2 is used along with the previous HCO_3^- to estimate the current pH using the Henderson-Hasselbalch equation. A capnography-free linear regression model by Baudin et. al[10] using $PetCO_2$, FiO_2, and mean airway pressure (MnAwP) was also tested on the validation dataset for comparison.

Performance comparison

Performance was evaluated by the 95% percentile of absolute error, or the worst 5% of samples. Performance of samples in separate pH ranges was reported. The percentage of samples with absolute error under 0.04 pH unit or 5 mmHg PCO_2 was calculated, following the CLIA gold standard for blood gas[19].

RESULTS

	Derivation dataset	Validation dataset
Final # of patients	1883	286
Final # of BG samples	12344	4030
Age (months)	50 ± 73	40 ± 62
CTICU **PICU***	60% of patients 41% of patients†	57% of patients 43% of patients
Metabolic acidosis‡ **Respiratory acidosis‡** **Metabolic alkalosis‡** **Respiratory alkalosis‡** **Mixed‡**	14% 28% 7% 6% 44%	10% 22% 17% 10% 41%
PARDS‡	**12%**	**9%**
Before resampling	**11%** pH < 7.3 **71%** 7.3 ≤ pH < 7.45 **18%** pH ≥ 7.45	**7%** pH ≤ 7.3 **71%** 7.3 ≤ pH < 7.45 **22%** pH ≥ 7.45
Post resampling and processing	**17%** pH <7.3 **54%** 7.3 ≤ pH < 7.45 **29%** pH ≥ 7.45	NA§
Blood gas type	**90%** arterial **10%** capillary	
Female	44%	

Table 2 *Summary of derivation and validation datasets. Numbers for post-processed data are shown, except for pH range data. *PICU: Pediatric (multidisciplinary, medical-surgical) ICU. †Patients may have stayed in both ICUs. ‡Definition of PARDS, respiratory and metabolic acidosis and alkalosis are discussed in Supplements (see Additional file bloodgas_supplementary_material.docx). §Validation dataset was not resampled.*

Cohort characteristics

Characteristics of the final derivation and validation cohorts are described in Table 2. The cohorts are representative of a general pediatric intensive care population. Sub-cohort criteria such as pediatric acute respiratory distress syndrome (PARDS) and respiratory acidosis are defined in Supplements (see Additional file bloodgas_supplementary_material.docx).

	Derivation		Validation	
pH 95% percentile (± pH units)	**Before abstention**	**After abstention**	**Before abstention**	**After abstention**
all	0.103	0.092	0.092	0.078
<7.3	0.114	0.096	0.096	0.083
7.3 – 7.45	0.103	0.094	0.094	0.075
≥7.45	0.089	0.076	0.076	0.083
PCO_2 95% percentile (± mmHg)	**Before abstention**	**After abstention**	**Before abstention**	**After abstention**
all	10.33	8.78	9.67	8.72
20 – 35	9.65	7.82	9.33	7.82
35 – 60	9.45	8.43	9.08	8.52
60 – 120	17.43	13.78	17.95	13.48

Table 3 *Blood gas estimation performance on derivation and validation datasets before and after abstention.*

Validation performance

Figure 3 plots estimates of pH and PCO_2 against laboratory values for the validation dataset, and estimation performance before and after abstention are shown in Table 3. Overall, estimations were within CLIA acceptable blood gas machine equivalents[19] in 74% of pH samples (±0.04 pH unit) and 80% of PCO_2 samples (±5 mmHg). Estimation accuracy was balanced across pH, especially after abstention. Using the Mann-Whitney U-test, the validation results outperformed AVDSf and Baudin[10] models (eTable 4) with statistical significance P-value <0.001%.

Prediction uncertainty and predictor importance

Estimations were not made when prediction uncertainty was above an acceptable threshold, as shown in the example in Fig. 3f. Also shown are the top three predictors ranked by importance and their measured values. The uncertainty threshold for abstention was obtained by examining the trade-off between performance and abstention rate on derivation samples, as shown in Fig. 4a. Abstaining using prediction uncertainty outperforms randomly abstaining the same percentage of samples, as shown in Fig. 4a, indicating that prediction uncertainty is a useful measure of

estimation confidence. After abstention results in Table 3 were obtained when the abstention rate was set to 25%, meaning that 75% of samples were estimated.

Prediction accuracy over time

Figure 4b examines the relationship between estimation accuracy and the time elapsed since the last BG. Samples were split into bins based on time lags. The 95% percentile remained under ±0.080 until up to an 8-hour time lag between the time of estimation and the previous BG.

While the most recent BG was used for modeling, an additional analysis examined using the first available BG for each patient for all subsequent estimations. The 95% percentile using the first available BG was ±0.124 pH unit, compared to ±0.078 pH unit when using the previous BG.

Safe classification of pH

As the predicted values are used to guide ventilator settings, erroneous predictions between pH ranges could be potentially dangerous for patients. Figure 4c examines whether the estimated pH range, spanned by the point estimate plus uncertainty range, cover the correct pH range. Overall, 85% of all estimations cover the correct pH range, while those in individual pH-ranges are above 70%.

Arterial and capillary blood gas

Estimations based on arterial BG were slightly more accurate than estimations based on capillary BG but not statistically different. The null hypothesis of no statistically significant difference in absolute estimation errors was not rejected by a Mann-Whitney U-test with P-value 0.07%.

Model visualization

Figure 2 depicts two example non-linear relationships learned between non-key predictors and the key predictor, previous pH. Contribution to estimated pH is color-coded onto the two-dimensional predictor value space, with yellow indicating higher estimated pH and purple indicating lower estimated pH. The left plot shows that a lower $etCO_2$ measurement contributes to a higher estimate of pH (as seen in the yellow color at the bottom of the plot), but how much higher varies based on the patient's previous pH. The right plot shows that the non-linear relationship between ΔSpO_2 and pH also varies depending on the key-predictor, previous pH.

DISCUSSION

This study demonstrates that noninvasive parameters routinely available on most clinical monitors and ventilators can be used to provide useful estimates of BG in all intubated patients without necessitating a new blood draw, for up to 8 hours. The model outperformed previous models while providing prediction uncertainty and predictor importance ranking, both of which can help users assess whether the model is likely to be accurate in a specific patient scenario. Built-in transparency of the model enables interpretation of estimation results, encouraging trust in adopting novel data-driven solutions for clinical practice.

Performance comparison to baseline models

The model estimated within CLIA acceptable blood gas machine equivalents[19] in 74% of pH samples (±0.04 pH unit) and 80% of PCO_2 samples (±5 mmHg). The model achieved better performance than previously reported models[10,11], especially in low-pH samples. Prediction

accuracy on validation data was comparable to that on derivation data, demonstrating that the model is generalizable to new data.

Correctly predicting the pH range

The pH ranges in this study were used by the ARDSNet studies[20]. While accuracy in the absolute value of a pH estimation is important, users may be more concerned with whether pH is estimated in the correct range. For example, it would be very detrimental for a pH of 7.15 (low) to be estimated as 7.45 (high) since the likely change in ventilator management would be rather different under the two clinical situations, whereas an inaccurate estimation of 7.25 (compared to 7.15) would result in a less impactful modification to treatment and the change suggested would be in the same direction as that for the lower pH of 7.15. We showed that the majority of estimated samples cover the correct pH range in Fig. 4c. Low-pH samples remain the most challenging samples to estimate but using prediction uncertainty results in 74% of low-pH samples falling in the correct range.

Prediction accuracy over time

Using older BGs for prediction was less accurate than using more recent BGs. This is likely due in part to variable changes in patient condition with time. Estimation accuracy decreased with longer time intervals between time of estimation and prior blood sample, but estimations remained accurate until up to 8 hours, suggesting that typically one may abstain from blood draws up to 8 hours from the previous BG.

Clinical application

The model utilizes readily available data sources in ventilated patients to provide continuous monitoring of BG through estimation. Estimations are made with prediction uncertainty highlighting inherent uncertainty in the model and preventing the display of potentially inaccurate predictions. Furthermore, the model displays top predictors and their values, which gives the user more context around the estimation.

One could argue that in current practice, clinicians are already able to 'guesstimate' the BG pH based on the same data, and that any estimation that does not achieve laboratory-level performance is not useful. We argue there are merits of the model even at the current performance level. First, the model provides automatic and continuous monitoring of BG pH without any human effort, saving time and mental calculation even if the estimation is not perceived as better than a 'guesstimation'. Second, ranked top predictors can illuminate patient measurements and changes that may not have occurred to bedside caregivers. Third, it can be a good reassurance model for clinicians who want to check that their 'guesstimate' matches trends from thousands of prior patients from whom the model was developed. Finally, estimation uncertainty is displayed, and clinicians can always make sure that the model does no harm by opting to obtain a BG.

There are several potential applications of the model. First, noninvasive estimates of pH can decrease the number of blood draws further, and recommend that users obtain blood draws when they are most necessary. Second, continuously available estimates may facilitate standardized assessment of ventilator support and adherence to ventilator protocols, particularly those promoting lung protective recommendations. This has been implemented in the management of PARDS at Children's Hospital Los Angeles without any clinical issues[1]. Clinicians can accept or reject the protocol's recommendation or obtain a blood draw if not confident about the prediction. In addition, the majority of BG for ventilated children in respiratory failure with

PARDS lie in a normal to high range where the model performs well. Finally, the model has potential applications for closed loop ventilation, and will likely improve existing algorithms which use the $PetCO_2$ directly.

Limitations and future directions

The model uses a recent BG under the assumption that the patient's respiratory and metabolic conditions have not drastically changed. Many external, contextual, and patient conditions are not available or captured at the time of estimation, so the final decision at the bedside must be left to the expertise of clinicians. One potential direction for improvement is obtaining a large dataset of BG, physiological measurements, and ventilation parameters along with full volumetric capnography for all patients. Volumetric capnography may provide additional information about patients' respiratory states and prognoses[22, 23] not present in $PetCO_2$. which could enable more accurate estimation of BG.

Lastly, although the model was generalized to unseen patients from the same center, it is unknown whether the model would generalize to other centers. This also requires validation on additional data.

CONCLUSIONS

We developed and validated a BG estimation model that is able to estimate pH and PCO_2 more accurately and transparently on a larger and more diverse population of patients. Performance in the low- and high-pH ranges were especially improved compared to previously published

models[10,11]. 74-80% of estimations fell within CLIA acceptable BG machine equivalents[19]: ±0.04 pH unit and ±5 mmHg PCO_2. Estimations were accurate when performed within 8 hours of the last blood gas.

The main application of the model is likely to be when patients are in a relatively stable clinical state. Hopefully, it will be a viable tool for avoiding blood draws and facilitating continuous BG monitoring leading to more lung protective practices as we currently understand ventilation and oxygenation management[24]. Ventilator decision support protocols based on measurements of arterial BG have proven useful in the management of adult respiratory failure[20]. Accurate noninvasive measurements of arterial or capillary PCO_2 with subsequent prediction of pH could allow more frequent ventilator changes to optimize lung and diaphragm protective ventilation without BG analysis, which would be particularly useful in pediatric practice where fewer arterial lines are used [25].

Abbreviations

ABPd: diastolic arterial blood pressure
ABPm: mean arterial blood pressure
ABPs: systolic arterial blood pressure
AVDSf : alveolar dead-space fraction
BG: blood gas
CLIA: Clinical Laboratory Improvement Amendments of 1988[19]
CTICU: Cardiothoracic intensive care unit
PICU: Pediatric (multidisciplinary, medical-surgical) intensive care unit
CV: cross-validation
FiO_2: fraction of inspired oxygen
HR: heart rate
NBPd: diastolic noninvasive blood pressure
NBPm: mean noninvasive blood pressure
NBPs: systolic noninvasive blood pressure
MnAwP: mean airway pressure
OSI: oxygen saturation index = $100 \times FiO_2 \times MAP/SpO_2$
OI: oxygen index = $100 \times FiO_2 \times MAP/PaO_2$
PCO_2: arterial or capillary CO_2 pressure
$PaCO_2$: arterial CO_2 pressure
$PcCO_2$: capillary CO_2 pressure
PARDS: pediatric acute respiratory distress syndrome
$PetCO_2$: exhaled end-tidal CO_2
PEEP: positive end expiratory pressure
PIP: peak inspiratory pressure
PF ratio: PaO_2/FiO_2
SpO_2: pulse oximetry saturation
SF ratio: SpO_2/FiO_2
TVexp: expiratory tidal volume
TVin: inspiratory tidal volume
% leak: % gas leak around endotracheal tube during respiratory cycle

Declerations

Ethics approval and consent to participate
The study was approved by the Institutional Review Board at Children's Hospital Los Angeles with a waiver of informed consent (CCI-09-00126 and CCI-09-00287).

Consent for publication
Not applicable.

Availability of data and materials
The datasets analysed during the current study are not publicly available due to privacy concerns but are available from the corresponding author on reasonable request.

Competing interests
CN serves as consultant to Philips Research.

Funding
Philips Research internal funding.

Authors' contributions
CN, RK, MX, and JD contributed to the conception and design of the study. JD and CN contributed to data processing, JD, MX, and BC contributed to analysis of data, and model design and training. JD and CN drafted the article, and all authors provided critical revision.

Acknowledgements
Not applicable.

References

1. Khemani RG, Markovitz BP, Curley MAQ. Characteristics of Children Intubated and Mechanically Ventilated in 16 PICUs. *Chest* 2009;136(3):765–771.

2. Santschi M, Jouvet P, Leclerc F, 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. O'Horo JC, Maki DG, Krupp AE, Safdar N. Arterial catheters as a source of bloodstream infection: a systematic review and meta-analysis. *Crit Care Med* 2014;42(6):1334–1339.

4. Rowan CM, Speicher RH, Hedlund T, Ahmed SS, Swigonski NL. Implementation of continuous capnography is associated with a decreased utilization of blood gases. *J Clin Med Res* 2015;7(2):71–75.

5. Riker JB, Haberman B. Expired gas monitoring by mass spectrometry in a respiratory intensive care unit. *Crit Care Med* 1976;4(5):223–229.

6. McAslan TC. Automated respiratory gas monitoring of critically injured patients. *Crit Care Med* 1976;4(5):255–260.

7. Niehoff J, DelGuercio C, LaMorte W, et al. Efficacy of pulse oximetry and capnometry in postoperative ventilatory weaning. *Crit Care Med* 1988;16(7):701–705.

8. Hatle L, Rokseth R. The arterial to end-expiratory carbon dioxide tension gradient in acute pulmonary embolism and other cardiopulmonary diseases. *Chest* 1974;66(4):352–357.

9. Sivan Y, Eldadah MK, Cheah T-E, Newth CJL. Estimation of arterial carbon dioxide by end-tidal and transcutaneous PCO2 measurements in ventilated children. *Pediatr Pulmonol* 1992;12(3):153–157.

10. Baudin F, Bourgoin P, Brossier D, et al. Noninvasive Estimation of Arterial CO2 From End-Tidal CO2 in Mechanically Ventilated Children: The GRAeDIENT Pilot Study. *Pediatr Crit Care Med* 2016;17(12):1117–1123.

11. Khemani RG, Celikkaya EB, Shelton CR, et al. Algorithms to estimate PaCO2 and pH using noninvasive parameters for children with hypoxemic respiratory failure. *Respir Care* 2014;59(8):1248–1257.

12. McSwain SD, Hamel DS, Smith PB, et al. End-tidal and arterial carbon dioxide measurements correlate across all levels of physiologic dead space. *Respir Care* 2010;55(3):288–293.

13. Bhalla AK, Belani S, Leung D, Newth CJL, Khemani RG. Higher Dead Space Is Associated With Increased Mortality in Critically Ill Children. *Crit Care Med* 2015;43(11):2439–2445.

14. American College of Physicians. Annals of Internal Medicine. The TRIPOD Statement:

Explanation and Elaboration. https://annals.org/aim/fullarticle/2088542/transparent-reporting-multivariableprediction-model-individual-prognosisdiagnosis-tripod-explanation.

15. Harrison AM, Lynch JM, Dean JM, Witte MK. Comparison of simultaneously obtained arterial and capillary blood gases in pediatric intensive care unit patients. Critical care medicine. 1997 Nov 1;25(11):1904-8.

16. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics.* 2006 Dec 1;7(1):91.

17. Lou Y, Caruana R, Gehrke J, Hooker G. Accurate intelligible models with pairwise interactions. InProceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining 2013 Aug 11 (pp. 623-631).

18. Meyer, J.S., Ingersoll, C.G., McDonald, L.L. and Boyce, M.S., 1986. Estimating uncertainty in population growth rates: jackknife vs. bootstrap techniques. *Ecology*, 67(5), pp.1156-1166.

19. US Department of Health and Human Services. Clinical laboratory improvement amendments of 1988. PL100-578.42USC§ 201(1988)

20. Brower RG, Matthay MA, Morris AH, Schoenfeld D, Thompson BT, Wheeler AP. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. *N Engl J Med* 2000;342(18):1301–1308.

21. Hotz JC, Bornstein D, Kohler K, et al. Real-Time Effort Driven Ventilator Management: A Pilot Study. Pediatric Critical Care Medicine 2020; Published Ahead of Print. doi:10.1097/pcc.0000000000002556

22. Jonson B. Volumetric Capnography for Noninvasive Monitoring of Acute Respiratory Distress Syndrome. Am J Respir Crit Care Med 2018;198(3):396–398.

23. Sinha P, Calfee CS, Beitler JR, et al. Physiological Analysis and Clinical Performance of the Ventilatory Ratio in Acute Respiratory Distress Syndrome. *Am J Respir Crit Care Med* 2018;:rccm.201804–0692OC.

24. Khemani RG, Hotz JC, Klein MJ, et al. A Phase II randomized controlled trial for lung and diaphragm protective ventilation (Real-time Effort Driven VENTilator management). Contemporary Clinical Trials 2020;88:105893. doi:10.1016/j.cct.2019.105893

25. Sward K, Newth C. Computerized Decision Support Systems for Mechanical Ventilation in Children. Journal of Pediatric Intensive Care 2016;05(03):095–100.

Additional files

Additional file '1.tiff': Figure 1
Additional file '2.tiff': Figure 2
Additional file '3.tiff': Figure 3
Additional file '4.tiff': Figure 4
Additional file 'bloodgas_supplementary_material.docx': Supplementary Material

Image Legends

Figure 1 *Block diagrams of derivation and validation cohort sizes and extraction steps.*

Figure 2 *Two examples of learned non-linear pairwise relationships between non-key predictors and the key predictor. The key predictor, previous pH, is shown with non-key predictors $etCO_2$ and ΔSpO_2. Contribution of each non-key and key predictor pair to the total estimated pH is color-coded, with yellow indicating higher estimations and purple indicating lower estimations.*

Figure 3 *Predicted blood gas (BG) pH and PCO_2 results on validation samples. Subplots **a** and **b** show the scatter plots of the model generated point-estimate and laboratory-derived pH and PCO_2, while **c** and **d** show Bland-Altman plots for these estimations. Subplot **e** shows a patient example where the estimate at time $t = 0$ is made accurately with low uncertainty, and **f** shows a patient example where the estimate at time $t = 0$ is abstained on the basis of high prediction uncertainty. In the scatter plots (**a** and **b**), the blue shaded regions are the 95% percentile for all samples. The three pH or PCO_2 regions are separated by vertical and horizontal dashed lines. In the Bland-Altman plots, the middle solid line shows mean predicted error, and the top and bottom dashed lines show ± 1.96 standard deviation.*

Figure 4 ***a*** *demonstrates that using prediction uncertainty to abstain on high-uncertainty samples improves estimation accuracy, while randomly abstaining the same percentage of samples provides no accuracy improvement.* ***b*** *compares the estimation performance between samples with different time lags, defined as the time passed since the last BG.* ***c*** *shows the percentage of estimations that fall in the correct range of pH after abstention.*

Figures

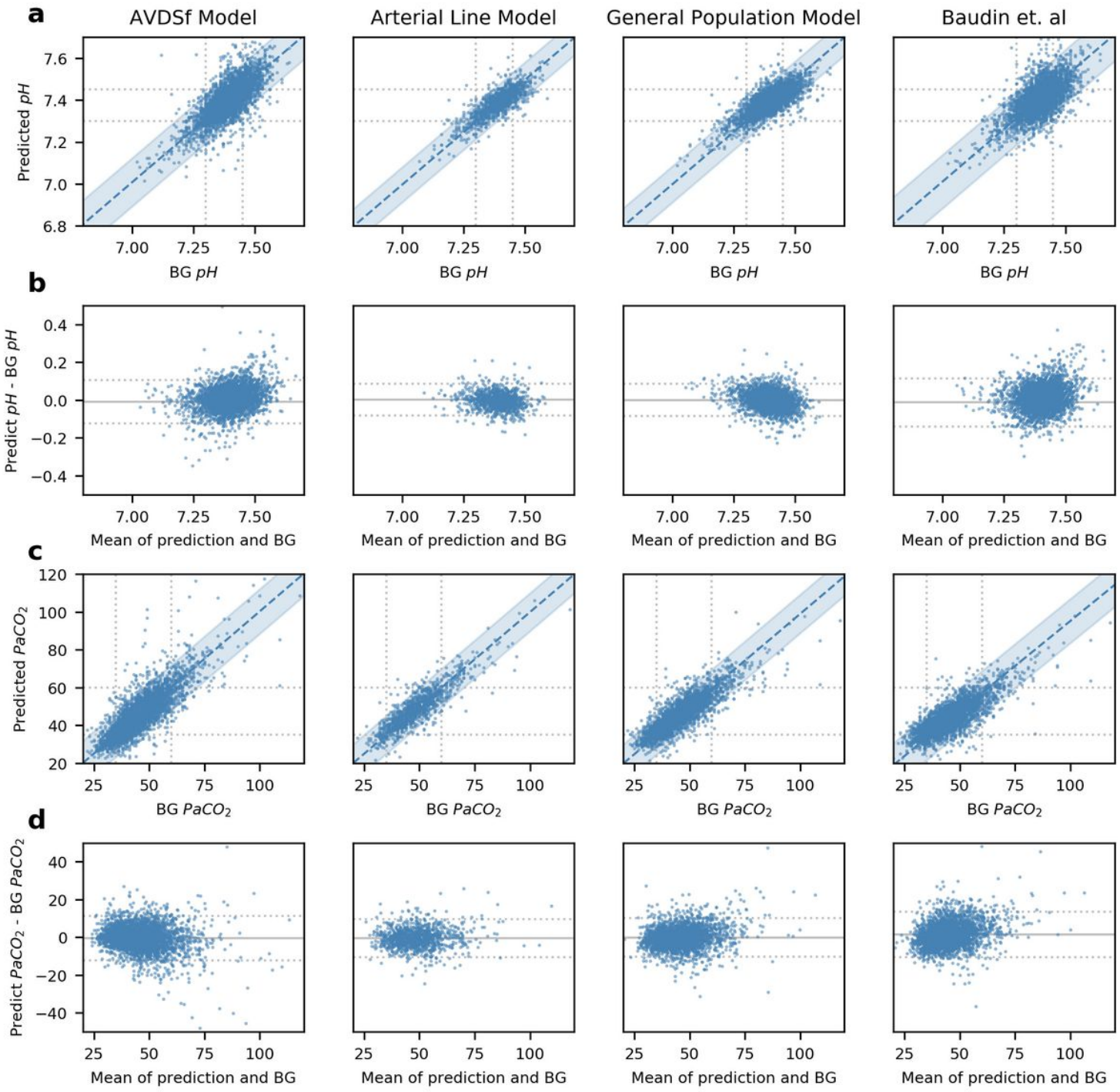

Figure 1

Block diagrams of derivation and validation cohort sizes and extraction steps.

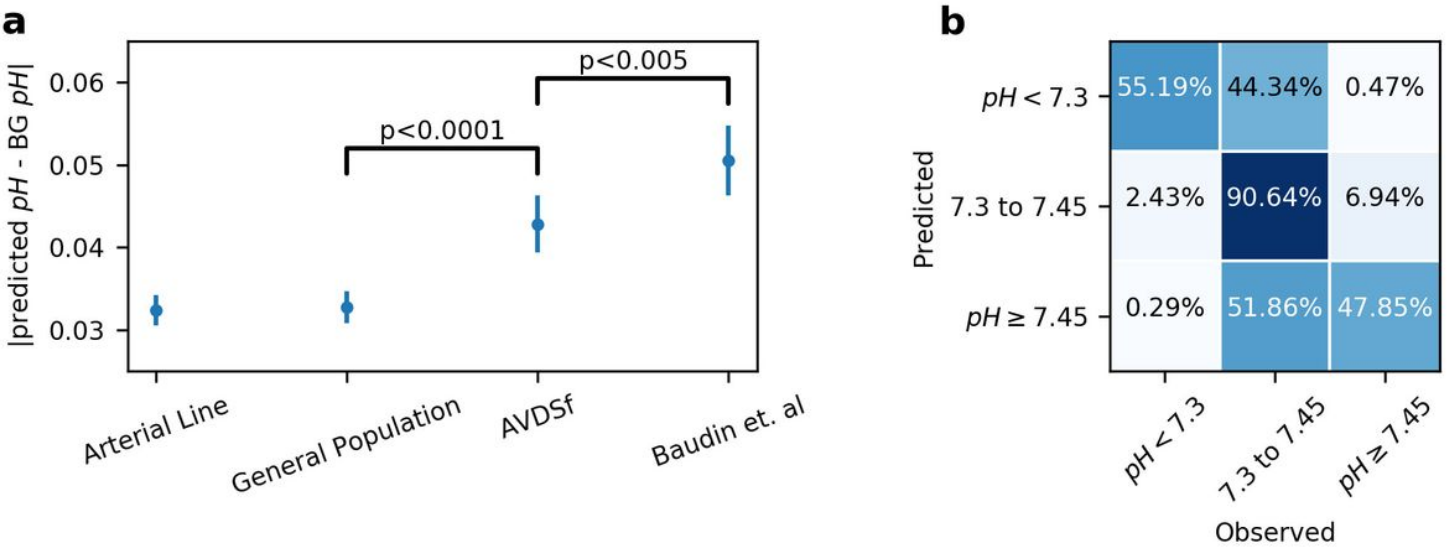

Figure 2

Two examples of learned non-linear pairwise relationships between non-key predictors and the key predictor. The key predictor, previous pH, is shown with non-key predictors etCO2 and ΔSpO2. Contribution of each non-key and key predictor pair to the total estimated pH is color-coded, with yellow indicating higher estimations and purple indicating lower estimations.

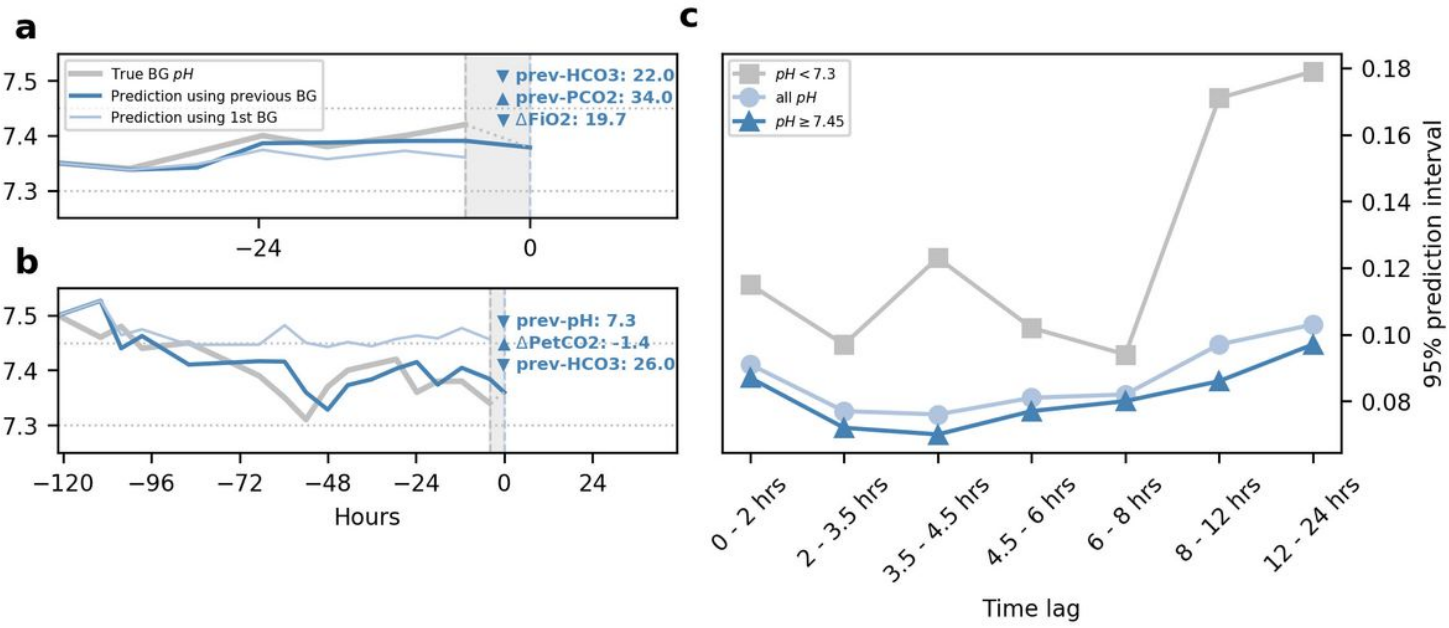

Figure 3

Predicted blood gas (BG) pH and PCO2 results on validation samples. Subplots a and b show the scatter plots of the model generated point-estimate and laboratory-derived pH and PCO2, while c and d show Bland-Altman plots for these estimations. Subplot e shows a patient example where the estimate at time t=0 is made accurately with low uncertainty, and f shows a patient example where the estimate at time

t=0 is abstained on the basis of high prediction uncertainty. In the scatter plots (a and b), the blue shaded regions are the 95% percentile for all samples. The three pH or PCO2 regions are separated by vertical and horizontal dashed lines. In the Bland-Altman plots, the middle solid line shows mean predicted error, and the top and bottom dashed lines show ±1.96 standard deviation.

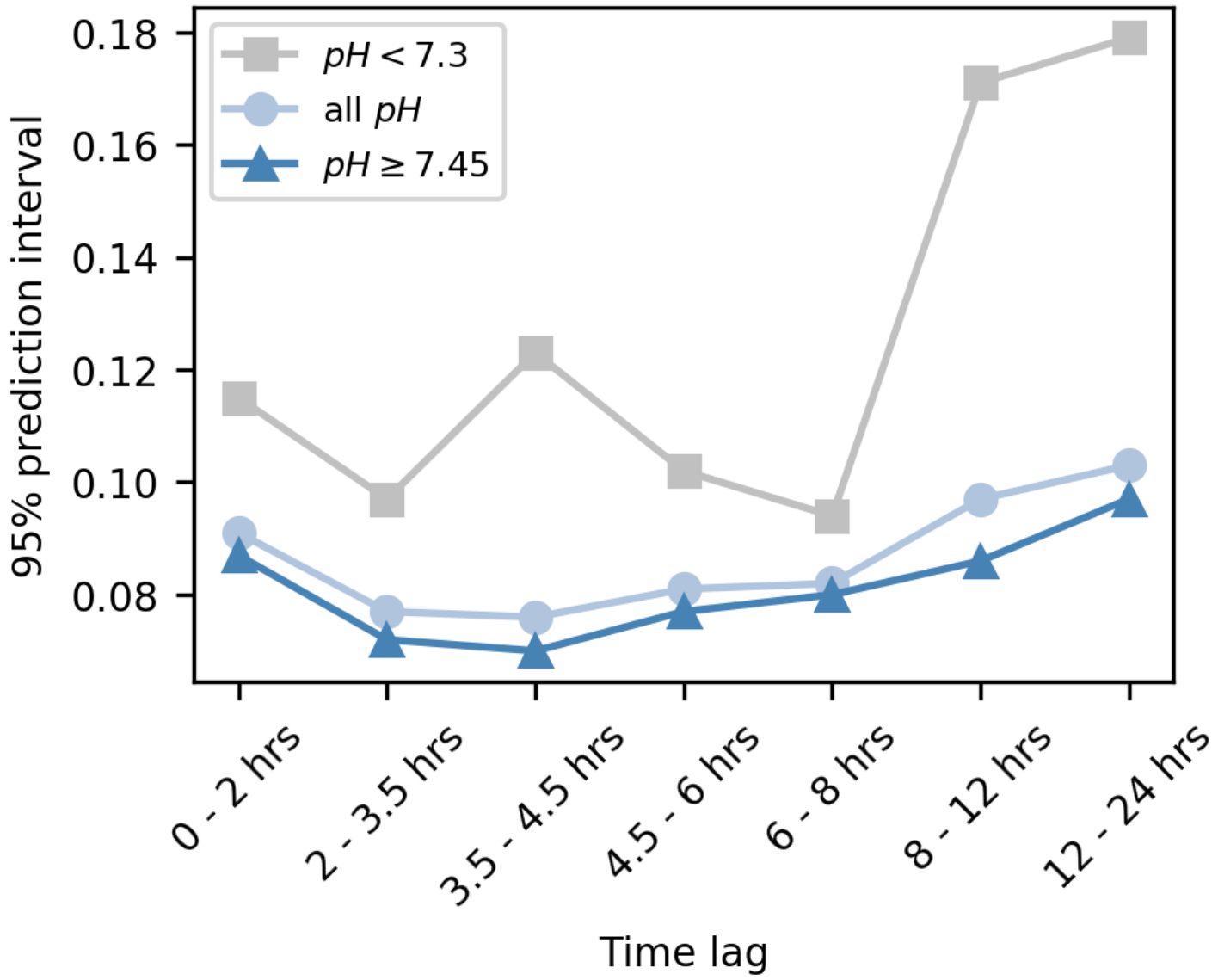

Figure 4

a demonstrates that using prediction uncertainty to abstain on high-uncertainty samples improves estimation accuracy, while randomly abstaining the same percentage of samples provides no accuracy improvement. b compares the estimation performance between samples with different time lags, defined as the time passed since the last BG. c shows the percentage of estimations that fall in the correct range of pH after abstention.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- bloodgassupplementarymaterial.docx