

# Proposal of a New Equation for Estimating Resting Energy Expenditure for Acute Kidney Injury Patients on Dialysis. A Machine Learning Approach.

daniela ponce (✉ [daniela.ponce@unesp.br](mailto:daniela.ponce@unesp.br))

Universidade Estadual Paulista Júlio de Mesquita Filho <https://orcid.org/0000-0002-6178-6938>

**Cassiana Regina de Goes**

Universidade Federal de Lavras Departamento de Ciencia dos Alimentos

**Luis Gustavo Modelli de Andrade**

Universidade Estadual Paulista Julio de Mesquita Filho Faculdade de Medicina Campus de Botucatu

---

## Research

**Keywords:** energy metabolism, resting energy expenditure, machine learning, acute kidney injury, sepsis, dialysis

**Posted Date:** July 13th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-37485/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on November 17th, 2020. See the published version at <https://doi.org/10.1186/s12986-020-00519-y>.

# Abstract

**Objective:** The objective of this study was, therefore, to develop a new predictive equation of resting energy expenditure (REE) for acute kidney injury patients (AKI) on dialysis.

**Material and methods:** A cross-sectional descriptive study has been carried out in 114 AKI patients on dialysis and mechanical ventilation consecutively selected, aged between 19 and 95 years. For the construction of the predictive model, 80% of the cases were randomly separated to train and 20% of unused cases for validation. Several machine learning models were tested in the training data: linear regression with Stepwise, rpart, support vector machine with radial kernel, generalized boosting machine, and random forest. The models were selected by 10-fold cross-validation and the performances evaluated by the root mean square error (RMSE).

**Results:** There were 364 indirect calorimetry (IC) measurements in 114 patients, mean age of  $60.65 \pm 16.9$  years and 68.4% were males. The average REE was  $2081 \pm 645$  kcal. REE was positively correlated with C-reactive protein, minute volume (MV), expiratory positive airway pressure, serum urea, body mass index and inversely with age. The principal variables included in the selected model were age, BMI, use of vasopressors, expiratory positive airway pressure, minute volume, C-reactive protein, temperature, and serum urea. The final r-value in the validation set was 0.69.

**Conclusion:** We propose a new predictive equation for estimating the REE for AKI patients on dialysis that use a non-linear approach with better performance than actual models.

## Introduction

Acute kidney injury (AKI) occurs in approximately 3–15% of hospitalized patients and can affect 30–50% of patients admitted to intensive care units (ICU). It is associated with extremely high mortality rates, ranging from 20 to 50% [1]. Previous observational studies reported that malnourished and hospitalized AKI patients have higher rates of morbidity and mortality than well-nourished patients [2, 3] and an association between cumulative caloric deficits and poor outcome in ICU patients [4, 5]. Accurate determination of energy needs is obviously important in critically ill patients because both over and underfeeding may be associated with complications and undesired consequences [6].

Determining energy requirements in critically ill patients via indirect calorimetry (IC) has long been considered the gold standard [7]. Limitations for using IC include time constraints, equipment availability, staffing, and cost. Therefore, many predictive equations exist for predicting resting energy expenditure (REE), but the accuracy of these equations for estimating caloric requirements for critically ill patients is not clear [7–13]. There are multiple equations for predicting resting energy expenditure (REE), but how accurate they are in severe AKI patients is not clear. Goes et al in a previous and recent study, evaluated if standard predictive nine different equations for energy expenditure accurately would reflect the energy requirements of critically ill, mechanically ventilated AKI patients. There were low precision, and poor agreement between measured and predicted REE by the Harris-Benedict (HB), Mifflin, Ireton-Jones, Penn

state, American College of Chest Physicians, and Faisy equations. HB without using the injury factor was the least precise (18% of precision). Modified Penn state equation had the best precision, although the precision rate was only 41%. As a conclusion, none of these equations accurately estimated measured REE in severe AKI patients on dialysis and most of them underestimated energy needs [14]. Recently more sophisticated models that used machine learning were applied in clinical practices resulting and better predict models [15]. These models can be applied to build predict equations aim to capture the non-linearity between variables resulting in better accuracy. These new models also have been applied to better predict resting energy expenditure [16].

The present study aimed, therefore, to develop a new predictive equation of REE for AKI patients on dialysis and mechanically ventilated using a machine learning approach.

## Material And Methods

A cross-sectional descriptive study has been carried out in the Dialysis Unit of Clinical Hospital from Botucatu Medical School. The project was approved by the center's ethical committee and the protocols used followed the criteria of the Helsinki Declaration (protocol 4383/2012). For the development of the predictive equation, 364 indirect calorimetry (IC) measurements in 114 AKI patients on dialysis were consecutively selected, aged between 19 and 95 years. Subjects with a body mass index (BMI) between 18–55 kg/m<sup>2</sup> were included, who gave their consent to participate in the study and agreed to comply with its standards. Inclusion criteria were patients admitted to the intensive care unit (ICU) with a diagnosis of AKI according to the KDIGO criteria [17] clinical symptoms suggestive of sepsis and acute tubular necrosis (ATN), need for renal replacement therapy (RRT) (stage 3), and mechanically ventilated. Exclusion criteria were patients with AKI of other etiologies, renal transplanted or those with chronic renal disease stages 4 and 5 (Glomerular Filtration Rate - GFR - < 30 ml/min estimated by the Modification of Diet in Renal Disease - MDRD) [18], a fraction of inspired oxygen (FiO<sub>2</sub>) greater than 0.60; Expiratory Positive Airway Pressure (PEEP) > 10 cm H<sub>2</sub>O; Maximum airway pressure > 60 cm H<sub>2</sub>O; stirring presence; use of neuromuscular blockers; air leakage into the ventilator circuit, around the endotracheal tube cuff, or from a bronchopleural fistula: because these factors lead to inaccuracies in the REE measurement by IC. IC was performed using the RMR Quark apparatus (Cosmed, Rome, Italy).

## Data analysis

For the construction of the predictive model, 80% of the cases were randomly separated. 20% of unused cases were separated for validation. The pre-processing was applied to train and test set: The continuous predictor variables were transformed using Box and Cox and after dividing by means (center) and standard deviation (scale). This transformation leads to a uniform scale (mean = 0, SD = 1) for all analyses so that they are comparable between analytical platforms. The missing variables were imputed by the median. Categorical variables were converted into dummy variables. In feature engineering, we tested the distribution of each continuous variable according to the outcome. We choose natural splines with 3 degrees of freedom for age and BMI aims to account the non-linearity. Afterward, linear and non-

linear models were tested in the training data. These models were: linear regression with stepwise selection, linear regression with regularization (glmnet), rpart, support vector machine with radial kernel (SVM), generalized boosting machine (GBM), extreme gradient boosting (XGBoost) and random forest. The optimization of hyperparameters was done in the train set with 10-fold cross-validation and the performance was evaluated by the root mean square error (RMSE). Therefore we evaluated the final performance of the best model in the test set by choosing the model with the lower RMSE (Fig. 01). The analysis was performed with R version 3.6.3, Vienna Austria, 2020 with caret package.

## Results

A total of 364 IC measures in 114 patients were evaluated. Mean age was  $60.65 \pm 16.9$  years and 68.4% were males. Diagnoses of sepsis and cardiovascular disease (CVD) accounted for 79.8% of the hospitalizations. AKI was the etiology associated with sepsis in most patients (81.6%). The BEE estimated by the Harris-Benedict formula averaged  $1,540 \pm 346$  Kcal; whereas the REE measured by the IC on was significantly higher  $2,081 \pm 645$  Kcal ( $27.7 \pm 10$  kcal/kg/day,  $p < 0.001$ ). Table 1 shows the clinical characteristics of the general population studied at the time when dialysis was indicated.

Table 1  
Clinical characteristics of the population with AKI on dialysis

<b>Patients Parameters (n = 114)</b>	
Age (years)	60.6 ± 16.9
Male gender (%)	78(68.4)
AKI (%)	
Sepsis-related	93 (81.6)
ischemic	13 (11.4)
Nephrotoxic	6(5.3)
Mixed	2 (1.8)
ATN-ISS	0,64 ± 0.18
Race (%)	
White 98 (86)	
Black	11 (9.6)
Weight (kg)	77,6 ± 22.4
Main diagnosis:	
CVD 35(30.7)	
Sepsis, severe sepsis, shock	56 (49.1)
Neoplasia	9(7,9)
Liver diseases	8 (7)
Trauma	6 (5.3)
Death	73 (63)
<b>Patients and IC Parameters (n = 364)</b>	
REE using HB (Kcal)	1540 ± 346
REE using IC (Kcal)	2081 ± 645
REE using IC (Kcal/Kg/d)	27.9 ± 10.4
VAD (mcg/Kg/	0.11

CVD: cardiovascular disease; AKI: acute kidney injury; ATN-ISS: Individual Severity Score in Acute Tubular Necrosis; REE: Resting Energy Expenditure, VAD: vasoactive drug; FIO2: Fraction of Inspired Oxygen; Mv: Minute volume; RR: respiratory rate; PEEP: expiratory positive airway pressure; Creat: serum creatinine; WBC: white blood-cell count; CRP: C-reactive protein;

<b>Patients Parameters (n = 114)</b>	
min)	(0.00-0.35)
VM	8.5 ± 2.8
FrEq. (resp/min)	17 ± 5.2
PEEP	6 ± 1.5
FIO <sub>2</sub>	38 ± 10.9
Temperature (°C)	37.6 ± 0.8
Urea (mg/dl)	157 ± 73
WBC (mm <sup>3</sup> )	17300(12300–25500)
PCR (mg/dl)	23.1(7.6–32.8)
CVD: cardiovascular disease; AKI: acute kidney injury; ATN-ISS: Individual Severity Score in Acute Tubular Necrosis; REE: Resting Energy Expenditure, VAD: vasoactive drug; FIO <sub>2</sub> : Fraction of Inspired Oxygen; Mv: Minute volume; RR: respiratory rate; PEEP: expiratory positive airway pressure; Creat: serum creatinine; WBC: white blood-cell count; CRP: C-reactive protein;	

The variables available to train the models of REE in the total population (n = 364) were summarized in attachment 01. In the exploratory data analysis, we plot the REE related to age and BMI (Fig. 02). We demonstrated a non-linear association between these predictors and the outcome. Then we applied natural splines with 3 degrees of freedom aim to capture the non-linearity (feature engineering).

We train different algorithms by 10-fold cross-validation in train set (80% of the data) and the hyperparameters of each model were tuned based in lower RMSE. After train tuned models, we apply resamples of the train data and selected the best model based in the lower RMSE (Fig. 3). Therefore, the final models were evaluated for performance in the test set (internal validation) aim to confirm the results of the train set (Table 02).

Table 02  
 Summary statistics of final tuned models in the test set (20% of the data). R<sup>2</sup> (R squared), and root mean square error (RMSE).

Models	R <sup>2</sup>	RMSE
Linear Model with Stepwise	0.28	569.24
Regularized Linear model	0.26	578.44
SVM	0.38	530.05
Random Forest	0.48	490.05
R part	0.36	566.61
GBM	0.36	535.88
XGBoost	0.45	499.75

The models that had a non-linear approach have a lower RMSE. The linear model may have had the advantage of simplicity and interpretability (attachment 02). The model with the best accuracy was selected by the lower RMSE: random forest. The variable importance of the final model (random forest) was plotted in Fig. 04.

Finally, we plotted a correlation between the predicted values of REE and observed REE in the test set (unseen data). The random forest model (best model) has a final r-value in this test set of 0.69 compared to an r value of 0.24 with the Harris-Benedict equation (Fig. 05).

## Discussion

Information about the energy expenditure assessment of patients with AKI is scarce in the literature. Our study found that the REE estimated by the Harris and Benedict formula [19] was significantly lower than that measured by IC. This finding corroborates the indication not to use this formula in critically-ill patients and in patients with AKI [11, 14, 20, 21] and the need to propose a new equation for AKI on dialysis. This study aimed to develop and validate predictive equations for REE in severe AKI patients using a machine learning approach. It was found that the models developed validly and significantly predict REE in these patients and according to several linear and non-linear algorithms. In the present study, REE was positively correlated with C-reactive protein, minute volume (MV), expiratory positive airway pressure, serum urea, body mass index and inversely with age (attachment 02). The principal variables included in the best model were age, BMI, use of vasopressors, expiratory positive airway pressure, minute volume, C-reactive protein, temperature, and serum urea. The final r-value in the validation set was 0.69. In the literature, there is no consensus regarding the procedures to be used for the validation of predictive models. Some authors do not suggest the use of determination and correlation coefficients for the validation of techniques or estimated variables [15]. Others consider that the Bland

Altman plot is likely to show a systematic proportion bias [19]. We used 10-fold cross-validation and select the model with lower RMSE. The performance of the best model was confirmed in a test set of 20% random selected unseen data (internal validation). The linear models had the advantage of simplicity to the model built and better interpretability but do not capture the non-linearity of the data. We confirm this with a very low accuracy of the Harris-Benedict equation in the test set. The use of natural splines to age and BMI predictors improves the linear models but does not reach the same performance as the non-linear models like boost trees or support vector machines [22–23]. The best model had a higher performance but in trade-off lower interpretability. In addition, the traditional model like Harris-Benedict [19] uses linear models that were easy to implement. Otherwise, the implementation of a model like the random forest has a necessity of a calculator. However, in present, we have the facility of computers or apps that may overcome this difficulty.

Another advantage of the machine learning approach is to demonstrate others predictor variables that influence the outcome by finding complex interactions. The principal factor that contributes to the variability of REE was the BMI and age that has been previous describe with traditional models [24–25]. The machine analysis in this study reveals other new contributed variables that predict de REE for example ventilator parameters and biochemical values. Otherwise, a simple linear model with this data set without pre-processing and feature engineering results in a model with lower accuracy and do not shown new interactions ( $R^2 = 0.21$ , data not shown).

Thus, this study presents the importance of estimating the REE in severe AKI patients, which is a determining factor in the assessment of nutrition status in critically ill patients. From a practical and application point of view, it is worth mentioning the capacity for technical evaluations to reduce the difference in REE between IC and estimated by others conventional formulas as HB, (ii) the use of these predictive equations for the assessment of passive and active trawling, contributing with relevant information for nutritionists and physicians. The present study has as main limitations the predictive models are not valid for no severe AKI patients and the Bland Altman model was not performed. Although we train models with a robust estimator with 10-fold cross-validation using RMSE as metric an approach regular used in machine learning analysis. We test the model in unseen data that may be considered an internal validation.

## **Conclusion**

We propose a new predictive equation for estimating the REE for AKI patients on dialysis that use the non-linear approach with better performance than actual models.

## **Declarations**

## **Ethics approval and consent to participate**



The project was approved by the center's ethical committee of Botucatu Medical School and the protocols used followed the criteria of the Helsinki Declaration (protocol 4383/2012).

### **Consent for publication**

Not applicable

### **Availability of data and materials**

All data generated or analysed during this study can be included in this published article as supplementary information files.

### **Competing interests**

The authors declare that they have no competing interests

## **Funding**

The study has no funding

## **Authors' contributions**

CRG and DP analyzed and interpreted the patient data regarding the acute kidney injury disease and disease and resting energy. DP and LGMA performed the a major contributor in writing the manuscript. LGMA performed the statistical analysis All authors read and approved the final manuscript.

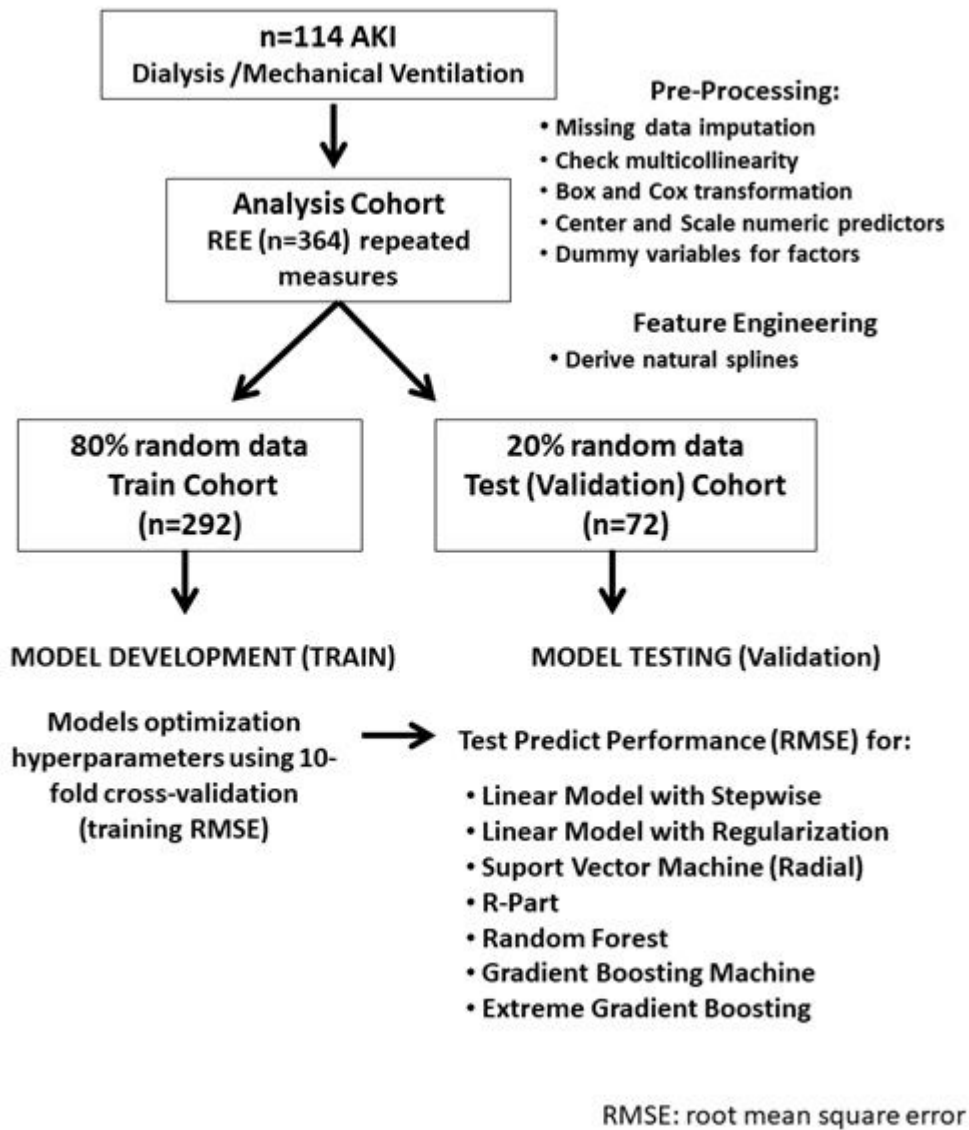
## **References**

1. VA/NIH Acute Renal Failure Trial Network  
Intensity of renal support in  
VA/NIH Acute Renal Failure Trial Network. Palevsky PM, Zhang JH, O'Connor TZ, Chertow GM, Crowley ST, et al. Intensity of renal support in.
2. critically ill patients with acute kidney injury. N Engl J Med. 2008;3(359):7e20.
3. Hanna J, Nichol A. Acute renal failure and the critically ill. Anaesth Intensive Care Med. 2012;1(13):166e70.
4. Dvir D, Cohen J, Singer P. Computerized energy balance and complications in critically ill patients: an observational study. Clin Nutr. 2006;25:37e44.
5. Fiaccadori E, Lombardi M, Leonardi S, Rotelli CF, Tortorella G, Borghetti A. Prevalence and clinical outcome associated with preexisting malnutrition in acute renal failure: a prospective cohort study. J Am Soc Nephrol. 1999;10:581e93.

6. McCarthy MS, Phipps SC. Special nutrition challenges: current approach to acute kidney injury. *Nutr Clin Pract*. 2014;29:56e62.
7. Berbel MN, Góes RC, Balbi AL, Ponce D. Nutritional parameters are associated with mortality in acute kidney injury. *Clinics*. 2014;69:476e82.
8. Schlein KM, Coulter SP. Best practices for determining resting energy expenditure in critically ill adults. *Nutr Clin Pract*. 2014;29:44e55.
9. Haugen HA, Chan L-N, Li F. Indirect calorimetry: a practical guide for clinicians. *Nutr Clin Pract*. 2007;22(4):377e88.
10. Singer P, Singer J. Clinical guide for the use of metabolic carts: indirect calorimetry-no longer the orphan of energy estimation. *Nutr Clin Pract*. 2016.;31:30e8.
12. Faisy C, Guerot E, Diehl JL, Labrousse J, Fagon JY. Assessment of resting energy expenditure in mechanically ventilated patients. *Am J Clin Nutr*. 2003;78:241–9.
13. Frankenfield D, Hise M, Malone A, Russell M, Gradwell E, Compher C. Prediction of resting metabolic rate in critically ill adult patients: results of a systematic review of the evidence. *J Am Diet Assoc*. 2007;107:1552–61.
14. Frankenfield DC, Coleman A, Alam S, Cooney RN. Analysis of estimation methods for resting metabolic rate in critically ill adults. *JPEN J Parenter Enteral Nutr*. 2009;33:27–36.
15. MacDonald A, Hildebrandt L. Comparison of formulaic equations to determine energy expenditure in the critically ill patient. *Nutrition*. 2003;19:233–9.
16. de Góes CR, Berbel-Bufarah MN, Sanches AC, Xavier PS, Balbi AL, Ponce D. Poor Agreement between Predictive Equations of Energy Expenditure and Measured Energy Expenditure in Critically Ill Acute Kidney Injury Patients. *Ann Nutr Metab*. 2016;68:276–84. DOI:<http://dx.doi.org/10.1159/000446708>.
17. [10.1371/journal.pone.0228597](https://doi.org/10.1371/journal.pone.0228597)  
Costa SD, de Andrade LGM, Barroso FVC, Oliveira CMC, Daher EF, Fernandes PFCBC, et al. The impact of deceased donor maintenance on delayed kidney allograft function: A machine learning analysis. *PLoS One*. 2020 Feb 6;15(2):e0228597. doi: [10.1371/journal.pone.0228597](https://doi.org/10.1371/journal.pone.0228597). eCollection 2020.
18. Armbruster M, Rist M, Seifert S, Frommherz L, Weinert C, Mack C, et al. Metabolite profiles evaluated, according to sex, do not predict resting energy expenditure and lean body mass in healthy non-obese subjects. *Eur J Nutr*. 2019 Sep;58(6):2207–17. doi:[10.1007/s00394-018-1767-1](https://doi.org/10.1007/s00394-018-1767-1). Epub 2018 Jul 4.
19. Kidney Disease. Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO clinical practice guideline for acute kidney injury. *Kidney*. 2012;2:1e138.
20. [10.1007/s00394-018-1767-1](https://doi.org/10.1007/s00394-018-1767-1)  
Int S. 2012;2:1e138.
21. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Modification of Diet in*

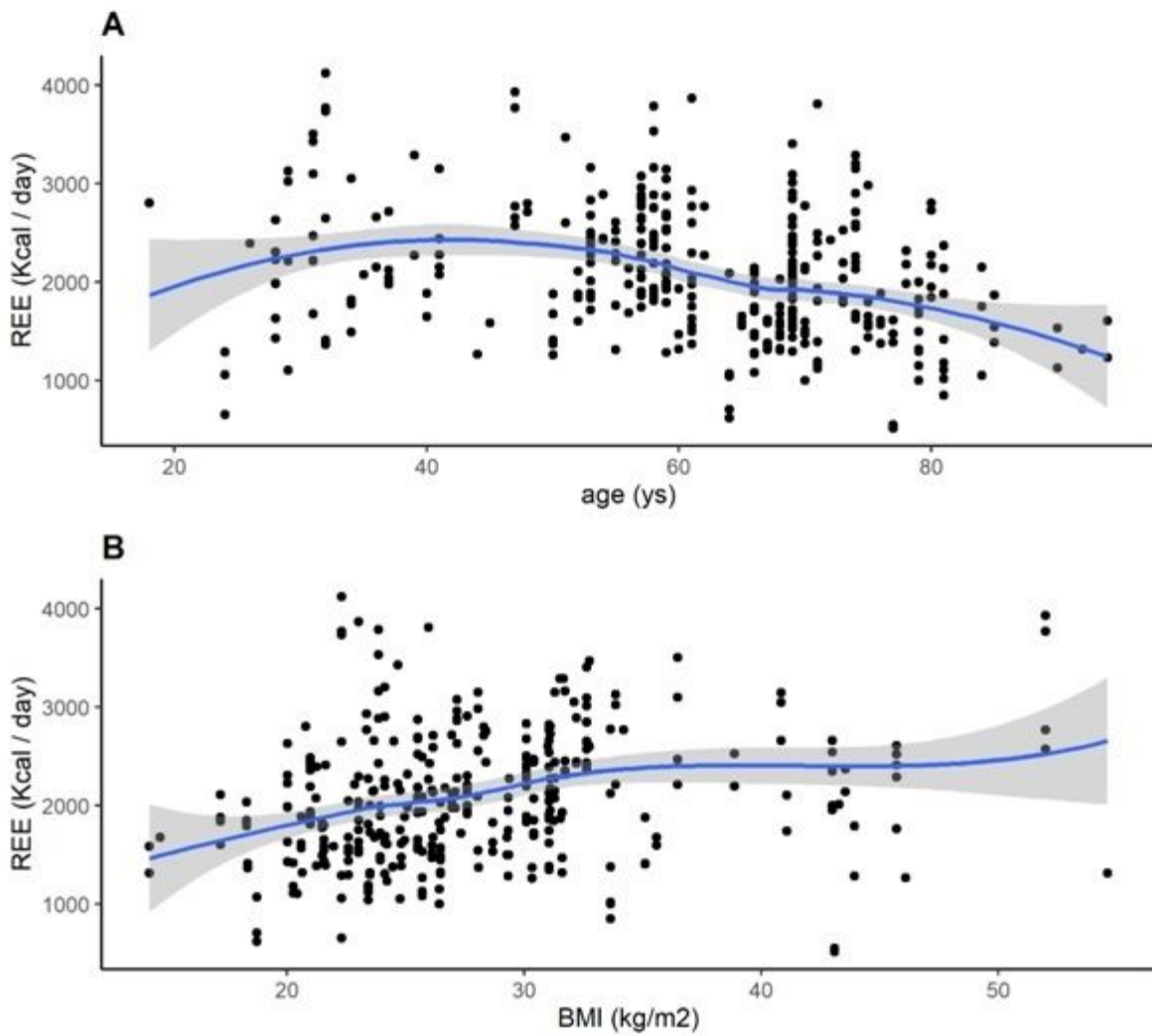
- Renal Disease Study Group. *Ann Intern Med.* 1999;16(130):461e70.
22. Harris JA, Benedict FG. A Biometric Study of Human Basal Metabolism. *Proc Natl Acad Sci U S A.* 1918;4:370–3.
23. 10.1016/j.jcrc.2011.07.084  
Kross EK, Sena M, Schmidt K, Stapleton RD. A comparison of predictive equations of energy expenditure and measured energy expenditure in critically ill patients. *J Crit Care* 2012;27:321. e5-12. DOI: <http://dx.doi.org/10.1016/j.jcrc.2011.07.084>.
24. <http://dx.doi.org/10.1007/s00134-014-3218-7>  
Hickmann CE, Roeseler J, Castanares-Zapatero D, Herrera EI, Mongodin A, Laterre PF. Energy expenditure in the critically ill performing early physical therapy. *Intensive Care Med* 2014;40:548 – 55. PMID: 24477456 DOI: <http://dx.doi.org/10.1007/s00134-014-3218-7>.
25. Gravesteyn BY, Nieboer D, Ercole A, Lingsma HF, Nelson D, van Calster B. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury.
26. 10.1371/journal.pone.0174944  
Weng SF, Reys J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One.* 2017 Apr 4;12(4):e0174944. doi: 10.1371/journal.pone.0174944. eCollection 2017.
27. 10.1016/j.jclinepi.2020.03.005  
Astrup A, Buemann B, Christensen NJ, Madsen J, Gluud C, Bennett P, Svenstrup B. The contribution of body composition, substrates, and hormones to the variability in energy expenditure and substrate utilization in premenopausal women. *J Clin Endocrinol Metab.* 1992;74(2):279–286. *J Clin Epidemiol.* 2020 Mar 19. pii: S0895-4356(19)30875-3. doi: 10.1016/j.jclinepi.2020.03.005. [Epub ahead of print].
28. Cunningham JJ. Body composition as a determinant of energy expenditure: a synthetic review and a proposed general prediction equation. *Am J Clin Nutr.* 1991;54(6):963–9. doi:10.1093/ajcn/54.6.963.

## Figures



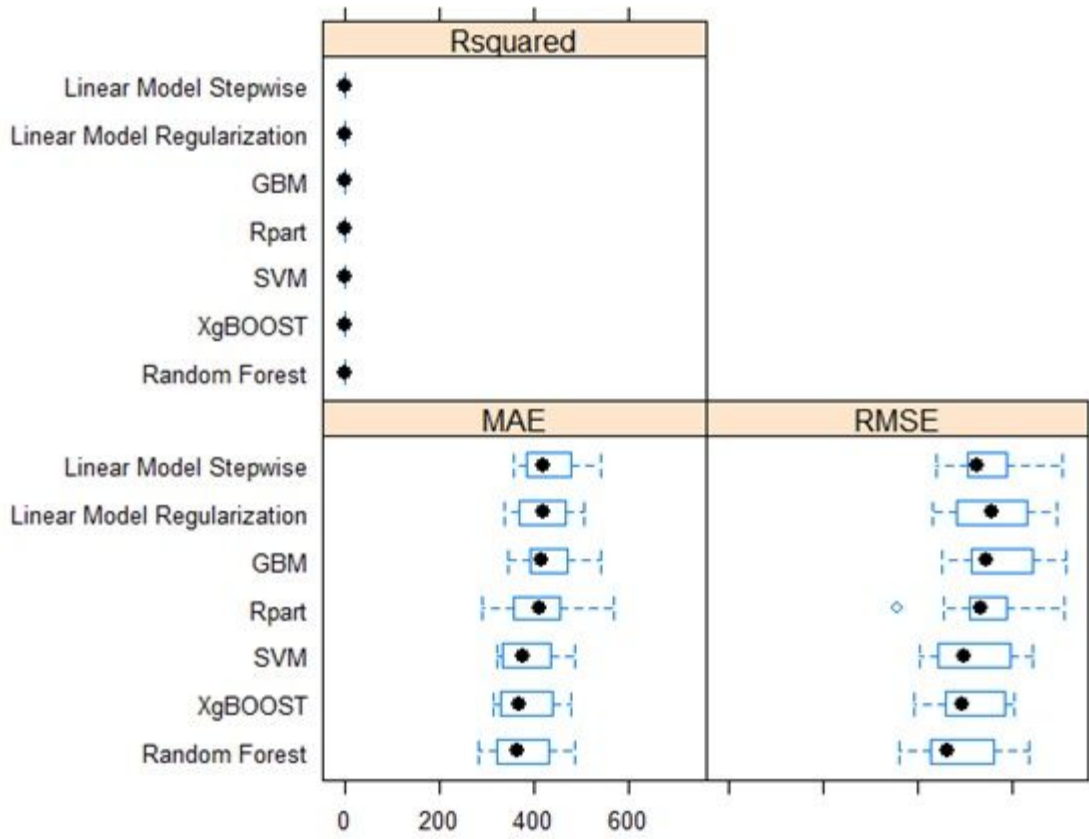
**Figure 1**

Flowchart of study design and description of the machine learning approach.



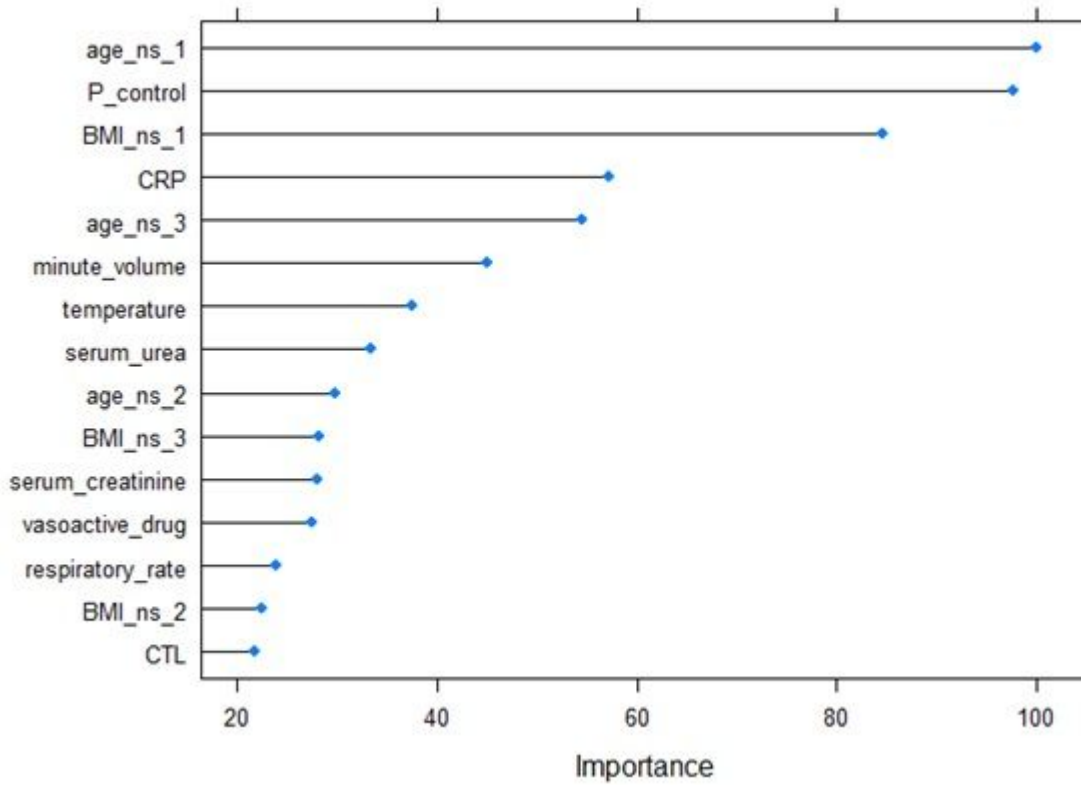
**Figure 2**

A plot of age-related to REE (A) and BMI related to REE (B). The blue line is a smooth that was fitting using a polynomial regression.



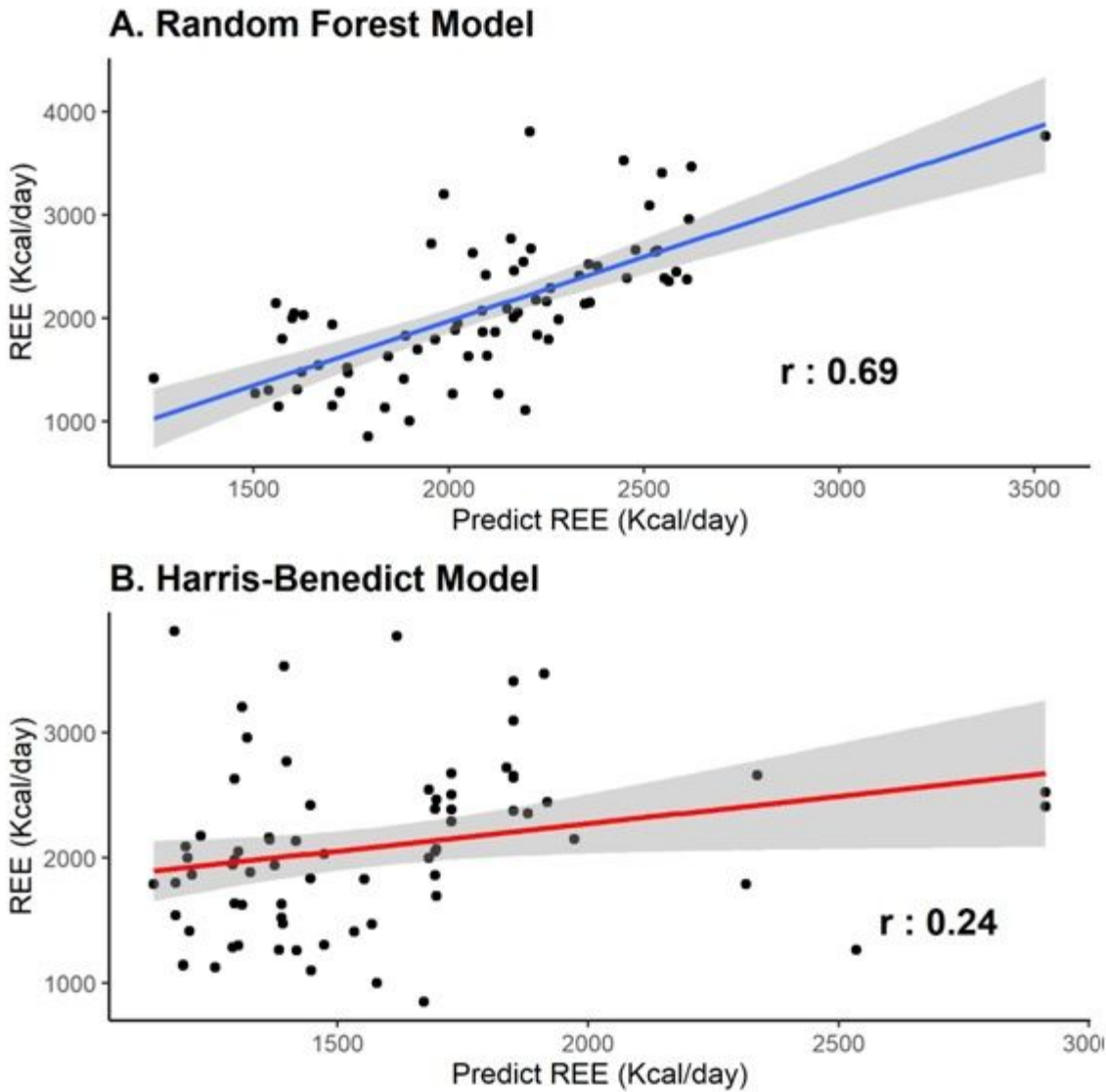
**Figure 3**

Summary statistics of tuned train models in resamples of the train set. The box represents mean, 25-75 percentile and min-max values. The statistics were: R squared; mean absolute error (MAE) and root mean square error (RMSE).



**Figure 4**

Variable importance in the random forest model. The importance was reported on a normalized scale.



**Figure 5**

A correlation plot between the predicted and observed REE in the test set (20% of unseen data). A: Predicted REE by random forest model and B: Predicted REE by Harris-Benedict equation. We had an r-value of 0.69 with a random forest model (new model) and an r-value of 0.24 with Harris-Benedict.