

RESEARCH ARTICLE

A New Data Analysis System to Quantify Associations between Biochemical Parameters of Chronic Kidney Disease-Mineral Bone Disease

Mariano Rodriguez¹, M. Dolores Salmeron¹, Alejandro Martin-Malo^{1*}, Carlo Barbieri², Flavio Mari², Rafael I. Molina¹, Pedro Costa³, Pedro Aljama¹

1 Nephrology Service, Hospital Reina Sofia, IMIBIC, University of Cordoba, Cordoba, Spain, **2** Fresenius Medical Care, Bad Homburg, Germany, **3** Fresenius Medical Care, Madrid, Spain

* alejandro.martin.sspa@juntadeandalucia.es



OPEN ACCESS

Citation: Rodriguez M, Salmeron MD, Martin-Malo A, Barbieri C, Mari F, Molina RI, et al. (2016) A New Data Analysis System to Quantify Associations between Biochemical Parameters of Chronic Kidney Disease-Mineral Bone Disease. *PLoS ONE* 11(1): e0146801. doi:10.1371/journal.pone.0146801

Editor: Martin H de Borst, University Medical Center Groningen and University of Groningen, NETHERLANDS

Received: July 10, 2015

Accepted: December 21, 2015

Published: January 25, 2016

Copyright: © 2016 Rodriguez et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by a restricted grant from Nephrologic Foundation of Cordoba (PI-0311-2014 Junta de Andalucia). Fresenius Medical Care provided support in the form of salaries for authors [CB, FM and PC], but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Abstract

Background

In hemodialysis patients, deviations from KDIGO recommended values of individual parameters, phosphate, calcium or parathyroid hormone (PTH), are associated with increased mortality. However, it is widely accepted that these parameters are not regulated independently of each other and that therapy aimed to correct one parameter often modifies the others. The aim of the present study is to quantify the degree of association between parameters of chronic kidney disease and mineral bone disease (CKD-MBD).

Methods

Data was extracted from a cohort of 1758 adult HD patients between January 2000 and June 2013 obtaining a total of 46.141 records (10 year follow-up). We used an advanced data analysis system called Random Forest (RF) which is based on self-learning procedure with similar axioms to those utilized for the development of artificial intelligence. This new approach is particularly useful when the variables analyzed are closely dependent to each other.

Results

The analysis revealed a strong association between PTH and phosphate that was superior to that of PTH and Calcium. The classical linear regression analysis between PTH and phosphate shows a correlation coefficient is 0.27, $p < 0.001$, the possibility to predict PTH changes from phosphate modification is marginal. Alternatively, RF assumes that changes in phosphate will cause modifications in other associated variables (calcium and others) that may also affect PTH values. Using RF the correlation coefficient between changes in serum PTH and phosphate is 0.77, $p < 0.001$; thus, the power of prediction is markedly increased. The effect of therapy on biochemical variables was also analyzed using this RF.

Competing Interests: The authors have the following interests: CB, FM and PC are employed by Fresenius Medical Care. MR has received honorarium for lectures and consultations from: Shire, Sanofi, Abbvie, Amgen, Vifor, Fresenius and Keryx. AMM has received honorarium for lectures from: Fresenius, Shire, Abbvie, Amgen and Bellco. PA has received grants from: Allexion, Astellas, Fresenius and Abbvie. Honorarium for lectures from: Astellas, Fresenius and Otsuka. MDS has received financial support from the Foundation for Biomedical Research of Cordoba (FIBICO) that belongs to the University Hospital Reina Sofia (Cordoba). There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

Conclusion

Our results suggest that the analysis of the complex interactions between mineral metabolism parameters in CKD-MBD may demand a more advanced data analysis system such as RF.

Introduction

In hemodialysis (HD) patients, increased mortality is in part explained by the presence of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). Deviations of serum concentration of phosphate (P), calcium (Ca) or parathyroid hormone (PTH) from the values recommended by KDIGO (Kidney Disease Improving Global Outcomes) [1,2] are associated with a negative outcome [3–7]. There are a number therapeutic strategies aimed to correct the concentration of these parameters; certainly the rate of success in controlling these parameters is variable [8–12].

The regulation of these three parameters, Ca, P and PTH are not independent from each other [9]. A strategy designed to change and correct the value of one of these parameters may be associated with a divergent effect in one or two of the remaining parameters. Based on our understanding on the regulation of mineral metabolism, one could predict that the modification of a single parameter should be followed by a change in another parameter, which in turn is conditioned by the third parameter. Furthermore, the interrelationship among these three parameters is likely to be non-linear. In a given patient, excessive administration of calcium based phosphate binders may reduce serum P level but it may also increase serum Ca and reduce PTH [13]. Thus final result is that in a population of hemodialysis (HD) patients the concentration of these three parameters is the result of both an abnormal mineral metabolism and the treatment used to correct these parameters.

Classical statistical methods may not be optimal for the analysis of non-linear associations among variables simultaneously affected by non-trivial feedback loops. Non-trivial feedback loops refers to a situation in which one variable (such as PTH) causes a variation in a second variable (i.e phosphate) which in turn causes a variation on a third variable (i.e calcium) that can modifies the first variable and so on. And, these variations are not linear. In such a case the use of machine learning techniques can overcome these difficulties [14]. If a sufficiently large amount of data is provided, machine learning techniques (like Random Forest) are capable to generate robust mathematical models that codify relationships among variables [14, 15]. The reliability of these relationships is based on the fact that these relationships emerge directly from the data presented with no *a priori* assumptions [14].

The availability of large datasets is critical for machine learning techniques to work properly [16,17]. Therefore, in medical field, the interest in machine learning approaches is progressively growing due to the availability of electronic health records [15]. In the present study, the data analysis has been performed using a machine learning algorithm called Random Forest (RF) [16] toward a predictive analytic approach. To our knowledge this approach in CKD-MBD context is highly innovative. Nevertheless a few studies have been published in the field of nephrology showing the usefulness of RF to predict the risk of diabetic kidney disease [18], to identify biomarkers that predict kidney transplant outcome [19] and to analyze mRNAs in urine samples of kidney transplant recipients [20].

The objective of the present work was to analyze the complex interrelationships between serum concentrations of Ca, P and PTH in HD patients using the machine learning technique RF for data analysis. This study was not designed to investigate new mechanisms and factors involved in CKD-MBD.

Methods

Dataset Description

Data was extracted from a cohort of 1758 adult HD patients between January 1, 2000 and June 1, 2013 in the area of Cordoba. Patients belonged to 7 hemodialysis centers; 3 of them dependent on the Andalusian Health System and the remaining 4 centers belonging to the company Fresenius Medical Care. For each patient, demographic, clinical and monthly monitored biochemical characteristics were considered as well as CKD-MBD treatment. Only laboratory tests containing PTH (measured with the same assay), P and, Ca concentrations, were included in the analysis. In addition to the serum concentration of Ca, P, PTH and alkaline phosphatase, the following set of variables were selected: age, gender, dialysis vintage. Time was entered as another variable together with the corresponding biochemical values and treatment of that specific date. The type of treatment based on the administration of nutritional Vitamin D, Calcitriol, Paricalcitol, Calcimimetics, Calcium and non-calcium based phosphate binders was done following the demands of each patient. In general we followed the KDIGO guidelines for CKD-MBD [2]. Related drugs treatment were associated to 8 binary variables representing whether a patient received a specific drug type during the 30 days before the blood sample for monthly lab was taken.

Initially data contains 46,141 measures of biochemical parameters. Only patients with complete information (Ca, P, PTH, age, gender, dialysis vintage, and type of treatment) were included in the analysis ([S1 Dataset](#)).

The study was approved by the Ethics Committee Institutional Review (Comité Ético de Investigación del Hospital Universitario Reina Sofía). Written informed consent was given by patients to be included in the Nephrology Department Data Base; all records / information was anonymized and de-identified prior to analysis.

Random Forest for Data Analysis

The information obtained through the Random Forest (RF) analysis is based on a process of learning. A brief description of the RF procedure will uncover the advantages of this data analysis ([Fig 1](#)). Decision trees are constructed based on the values of many variables from many patients. Having enough data allows the construction of a large number of decision trees; as many as needed to obtain reliable predictions. In the present study the number of 100 trees was selected for generation of models. The output of RF is the average obtained from all different trees.

RF regression models for the prediction of Ca, PTH and P serum levels were derived.

In the model for prediction of Ca (the output or dependent variable), the rest of variables were used as input variables. The same procedure was applied to obtain models for the prediction of PTH and P. The RF model associates to every possible combination of input variables (e.g., PTH, phosphate, age, use of Calcimimetics, etc.) the prediction of a value for one output variable (e.g., calcium level). Seventy percent of the data is randomly selected to obtain models (training set), the remaining 30% is used as testing set to evaluate the prediction accuracy on un-seen examples (testing set). Accuracy is measured as the mean absolute error and the Pearson's correlation coefficient calculated considering the estimated values (model output) and the real values of Ca, PTH and P serum concentration. If the model is able to accurately predict the output variable, it can be used to predict the effect of changing the values of input parameters (e.g., taking or not a drug) on the output variable. It is possible to calculate the relative relevance of each input variable in the prediction of a model: this is called feature importance analysis.

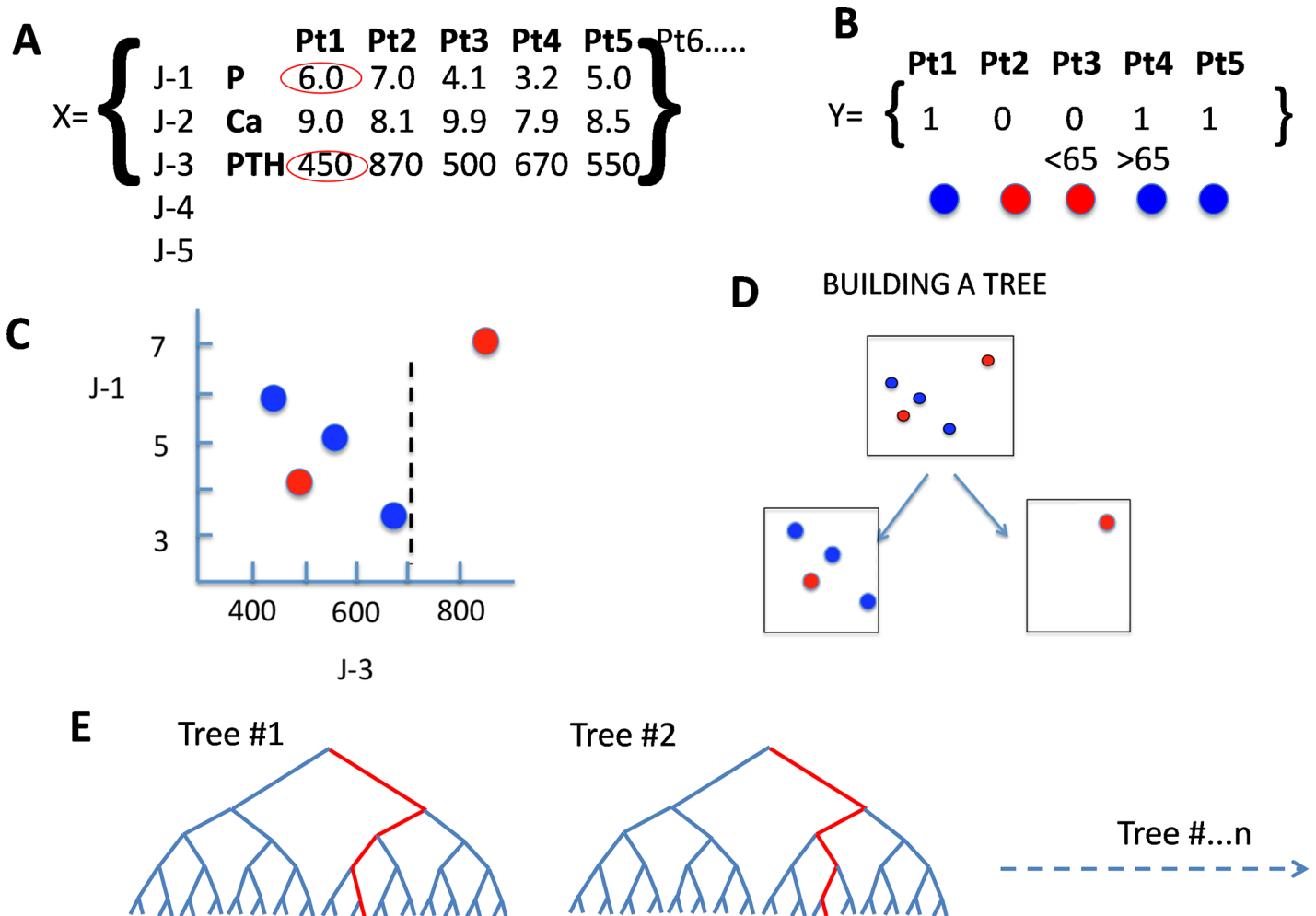


Fig 1. Brief Description Random Forest. (A) Dataset. (B) Outcome variable; (C and D) generation of decision trees. (E) Decision Trees. (A) The dataset “X” includes all variables (J-1, J-2, ..., J-n) such as serum levels of P, Ca, PTH, alkaline phosphatase, etc. The values of these variables are obtained from all patients included in the study: Pt1, Pt2, Pt3 up to Ptn. A first step is to construct “intelligent” decision trees based on the available data. (B) Let’s construct a decision tree that predicts an output variable Y that in this case will be age. Open and closed circles represent patients with age < 65 and >65 respectively). Decision trees will be constructed based on the values of many variables from many patients. (C) A set of two variables, J1 (P) and J3 (PTH) are chosen at random from a subset of data that is also chosen at random out of the entire dataset. The values of P are plotted against the corresponding PTH; open and closed circles represent patients with age < 65 and >65 respectively. The best PTH value that discriminate reds from blues is 700. (D) The first discrimination of the decision tree is that with a PTH greater than 700 all (100%) of patients are less than 65 and 75% of patients with less than 700 are more than 65. The next step will be to separate each one of the two groups according to any other variable chosen at random that could be serum calcium or any other. (E) This process will be done using, large number patients and will be repeated many times in different subsets of data at random. A large number of decision trees are constructed; as many as needed to obtain reliable predictions. Further information may be obtained in Antonio Criminisi, Jamie Shotton, Duncan Robertson, and Ender Konukoglu (Anatomy Regression Forests for Efficient Detection and Localization in CT Studies B. Menze et al. (Eds.): MCV MICCAI 2010 Workshop, LNCS 6533, pp. 106–117, 2011. c_Springer-Verlag Berlin Heidelberg 2011).

doi:10.1371/journal.pone.0146801.g001

The analyses were performed using Matlab version 8.2 (R2013b). In particular TreeBagger algorithm of the Statistics Toolbox was used for RF computation and Regress algorithm was used for linear regressions computation and for Pearson’s correlation coefficient.

Results

RF models for the prediction of PTH, Ca, and P were computed and optimized. Table 1 shows the characteristics of the study population at the baseline (first available measurement of biochemical parameters).

Table 1. Baseline characteristics of patients included in the study.

DEMOGRAPHIC VARIABLES	
Male (%)	59.2
Age. Mean (SD)	59.8 (16.6)
Vintage, days. Mean (SD)	3300 (2560)
Fistula (%)	75.5
BIOCHEMICAL PARAMETERS. MEAN (SD)	
Calcium (mg/dL)	9.4 (0.9)
Phosphate (mg/dL)	4.7 (1.4)
PTH (pg/mL)	317 (320)
Alkaline Phosphatase (IU/L)	142 (126)
Potassium (mmol/L)	5.3 (0.9)
CKD-MBD RELATED DRUGS ADMINISTRATION(% of patients on the medication)	
D-Vitamin (%)	1.4%
Calcitriol (%)	15.6%
Paricalcitol (%)	13.1%
Calcium Binder (%)	42.4%
Magnesium Binder (%)	0.6%
Sevelamer Binder (%)	41.8%
Lantano Binder (%)	10.6%
Calcimimetics (%)	15.5%

CKD-MBD: Chronic kidney disease-Metabolic bone disease; SD: Standard deviation

doi:10.1371/journal.pone.0146801.t001

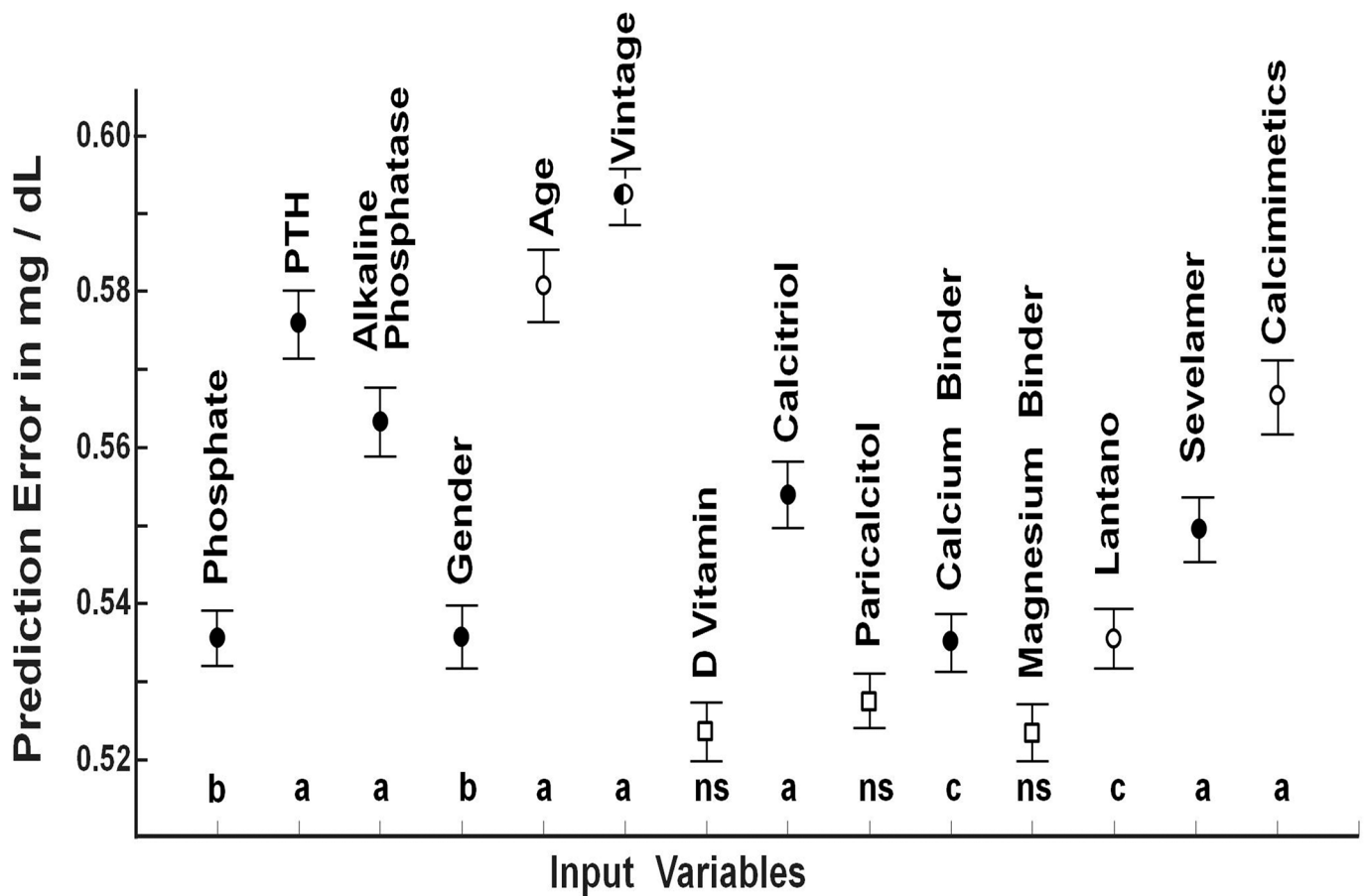
The values of output variables calculated from models obtained by RF were correlated with actual values obtained from the database. The correlation coefficients between the predicted output values and the real values were the following: Calcium model: $R_{Ca} = 0.60$ ($p < 0.001$); PTH model, $R_{PTH} = 0.67$ ($p < 0.001$); phosphate model, $R_p = 0.51$ ($p < 0.001$). These correlation coefficients are superior to those obtained using models derived from multivariate linear analysis: $R_{Ca} = 0.37$ ($p < 0.001$), $R_{PTH} = 0.47$ ($p < 0.001$) and $R_p = 0.37$ ($p < 0.001$).

Then, the RF models were explored through a feature importance analysis which evaluates the role of each input variable on the prediction of the output variables: Ca, PTH and P (Figs 2, 3 and 4). The relevance of each feature variable is assessed by dropping the variable from the model and evaluating to what extent the prediction error increases. If a variable is relevant, the model without that input variable will have a substantial increase in prediction error. The degree of increase on prediction error is a measure of the relevance of that input variable in the model. (Figs 2, 3 and 4).

Calcium model

Analysis of variable importance in the prediction of serum calcium concentration revealed that the most relevant variables are the serum concentrations of PTH, alkaline phosphatase, dialysis vintage and age (Fig 2). The vertical axis of Fig 2 depicts the \pm error value in the prediction of serum calcium concentration (mg/dl) that will be caused if specific input variable is removed from the mathematical model generated by RF. Of interest is the fact that, removing age and vintage from the model will cause an error in prediction of serum calcium of ± 0.57 and ± 0.58 mg/dl respectively. Serum Ca level is also closely and positively associated with serum concentration of PTH and alkaline phosphatase. The association between serum Ca and P levels is very modest.

Calcium: Features Analysis



p-Values: (a) $p < 0.001$; (b) $p < 0.01$; (c) $p < 0.05$; (ns) $p > 0.05$

Association: (●) Positive; (○) Negative; (◐) Positive if $< 8Y$ and negative if $> 8Y$; (□) no association

Fig 2. Prediction error for Serum Calcium concentration by each one of the variables included in the analysis. The vertical axis of Fig 2 depicts the error in prediction of serum calcium concentration (mg/dl) that will be caused if specific input variable is removed from the mathematical model generated by Random Forest. The magnitude of prediction error obtained with each variable is being compared with the rest of the variables. Only variables that show a significant effect in predicting values of calcium were included in this analysis.

doi:10.1371/journal.pone.0146801.g002

Calcimimetics and calcitriol appear to be associated to low and high calcium levels respectively. The administration of calcimimetics was associated to a decrease in serum calcium to a mean value of 9.0 mg/dl. Calcitriol increased the mean serum calcium to 9.8 mg/dl. The association between Paricalcitol and serum calcium level was modest. (Fig 2). None of the phosphate binders, except sevelamer, was associated to the value of serum calcium. Patients on sevelamer show serum calcium levels slightly higher than patients not taking sevelamer. This may reflect the fact that sevelamer was given to patients with high calcium to avoid calcium containing calcium binders.

The comparison of prediction error among variables included in the feature analysis for calcium is shown in Table 2.

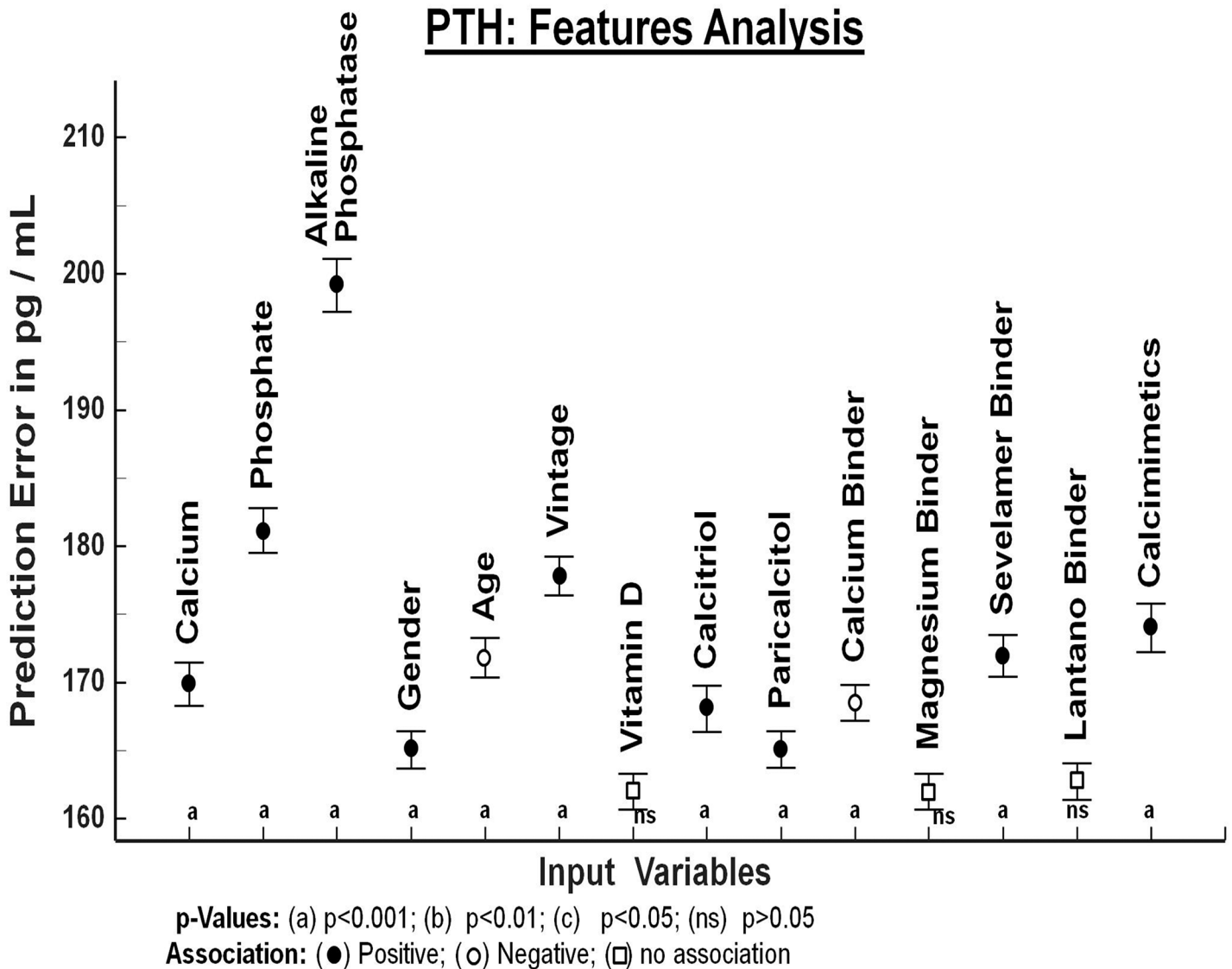


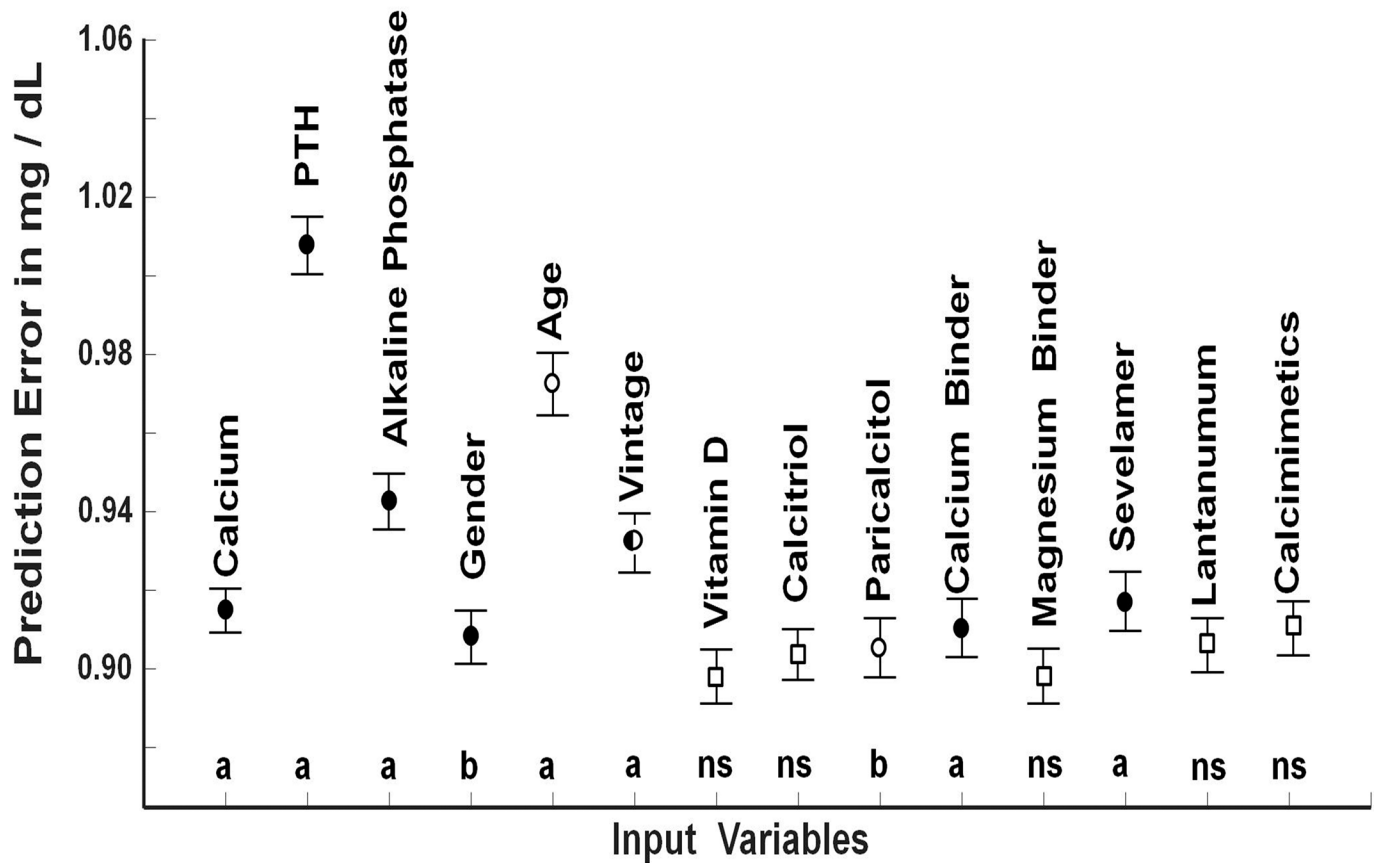
Fig 3. Prediction error for Serum PTH concentration by each one of the variables included in the analysis. The vertical axis of Fig 3 depicts the error in prediction of serum PTH concentration (pg/ml) that will be caused if specific input variable is removed from the mathematical model generated by Random Forest. Paricalcitol, Calcitriol and Calcimimetics are positively associated with PTH because this therapy is given to patients with hyperparathyroidism. The magnitude of prediction error obtained with each variable is being compared with the rest of the variables. Only variables that show a significant effect in predicting values of PTH were included in this analysis.

doi:10.1371/journal.pone.0146801.g003

PTH model

Variables that were associated with serum PTH concentration are shown in Fig 3. The vertical axis depicts the error in prediction of serum PTH concentration (pg/ml) that will be caused if specific input variable (horizontal axis) is removed from the mathematical model generated by RF. As expected, values of serum PTH concentration were closely associated to alkaline phosphatase. PTH was positively associated with dialysis vintage and negatively associated with age. There was also a noticeable positive association between serum concentrations of PTH and P that is clearly superior to the association PTH-Ca. In patients receiving calcimimetics the average serum PTH concentration was 423 pg/ml. PTH was also associated to the use of Calcitriol

Phosphate: Features Analysis



p-Values: (a) $p < 0.001$; (b) $p < 0.01$; (c) $p < 0.05$; (ns) $p > 0.05$

Association: (●) Positive; (○) Negative; (◐) Positive if $< 8Y$ and negative if $> 8Y$; (□) no association

Fig 4. Prediction error for Serum Phosphate concentration by each one of the variables included in the analysis. The vertical axis of Fig 4 depicts the error in prediction of serum phosphate concentration (mg/dl) that will be caused if specific input variable is removed from the mathematical model generated by Random Forest. Sevelamer and calcium binders are positively associated with serum phosphate concentration because patients with hyperphosphatemia receive phosphate binders. The magnitude of prediction error obtained with each variable is being compared with the rest of the variables. Only variables that show a significant effect in predicting values of phosphate were included in this analysis.

doi:10.1371/journal.pone.0146801.g004

and Paricalcitol; patients already on calcitriol had mean PTH values of 434 pg/ml while those on paricalcitol had a mean PTH of 359 pg/ml.

The comparison of prediction error among variables included in the feature analysis for PTH is shown in [Table 3](#).

Predicting the change in PTH resulting from a variation of serum Phosphate. The positive association between serum PTH concentration and serum phosphate levels was further analyzed using two different approaches. First, classical linear regression analysis shows that the serum concentrations of PTH and P are significantly correlated: $R = 0.27$, $p < 0.001$ ([Fig 5A](#)). Second, using the mathematical models derived from RF data analysis we predicted the change in PTH resulting from a variation in serum P concentration. A random variation was imposed to P (from -3 mg/dl to +3 mg/dl) and the difference between estimated PTH before

Table 2. Comparison of prediction error among variables included in the feature analysis for calcium.

	PTH	AP	G	Age	Vint.	CTR	Ca Bind.	Sev	CM
P	<0.001	<0.001	ns	<0.001	<0.001	<0.001	ns	<0.001	<0.001
PTH	-	<0.001	<0.001	ns	<0.001	<0.001	<0.001	<0.001	<0.001
AP	-	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	ns
G	-	-	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Age	-	-	-	-	<0.001	<0.001	<0.001	<0.001	<0.001
Vint	-	-	-	-	-	<0.001	<0.001	<0.001	<0.001
CTR	-	-	-	-	-	-	<0.001	Ns	<0.001
Ca Bind.	-	-	-	-	-	-	-	<0.001	<0.001
Sev.	-	-	-	-	-	-	-	-	<0.001

Abbreviations: (P) Phosphate; (AP) Alkaline Phosphatase, (G) Gender, (Vint) Vintage, (CTR) Calcitriol, (Ca Bind.) Calcium binders, (Sev.) Sevelamer. (CM) Calcimimetics

doi:10.1371/journal.pone.0146801.t002

and after P variation was computed (Fig 5B). This approach reveals a strong relationship between the changes in serum P and PTH: $R = 0.77$ ($p < 0.001$). According to these analyses a decrease in serum P of 2 mg/dl may be associated to a change in PTH of +50 to -250 pg/ml, as shown in Fig 5(B).

Phosphate model

Variables that were associated with serum P concentration are shown in Fig 4. Serum phosphate concentration was positively associated with serum PTH and negatively associated with age. Serum P level was associated to the use of phosphate binders: calcium binders, sevelamer and lanthanum without significant differences among the phosphate binders

The comparison of prediction error among variables included in the feature analysis for phosphate is shown in Table 4.

Discussion

The aim of the present study is to investigate the utility of the new data analysis system Random Forest to quantify associations between parameters of mineral metabolism in HD

Table 3. Comparison of prediction error among variables included in the feature analysis for PTH.

	P	AP	G	Age	Vint.	CTR	Paric	Ca Bind.	Sev	CM
Ca	<0.001	<0.001	<0.001	ns	<0.001	ns	<0.001	ns	ns	<0.001
P	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
AP	-	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
G	-	-	-	<0.001	<0.001	0.016	ns	ns	<0.001	<0.001
Age	-	-	-	-	<0.001	<0.001	<0.001	<0.001	ns	ns
Vint	-	-	-	-	-	<0.001	<0.001	<0.001	<0.001	<0.001
CTR	-	-	-	-	-	-	<0.001	ns	0.001	<0.001
Paric	-	-	-	-	-	-	-	0.001	<0.001	<0.001
Ca Bind	-	-	-	-	-	-	-	-	<0.001	<0.001
Sev	-	-	-	-	-	-	-	-	-	ns

Abbreviations: (P) Phosphate; (AP) Alkaline Phosphatase, (G) Gender, (Vint) Vintage, (CTR) Calcitriol, (Paric) Paricalcitol, (Ca Bind.) Calcium binders, (Sev.) Sevelamer. (CM) Calcimimetics, (Ca) Calcium

doi:10.1371/journal.pone.0146801.t003

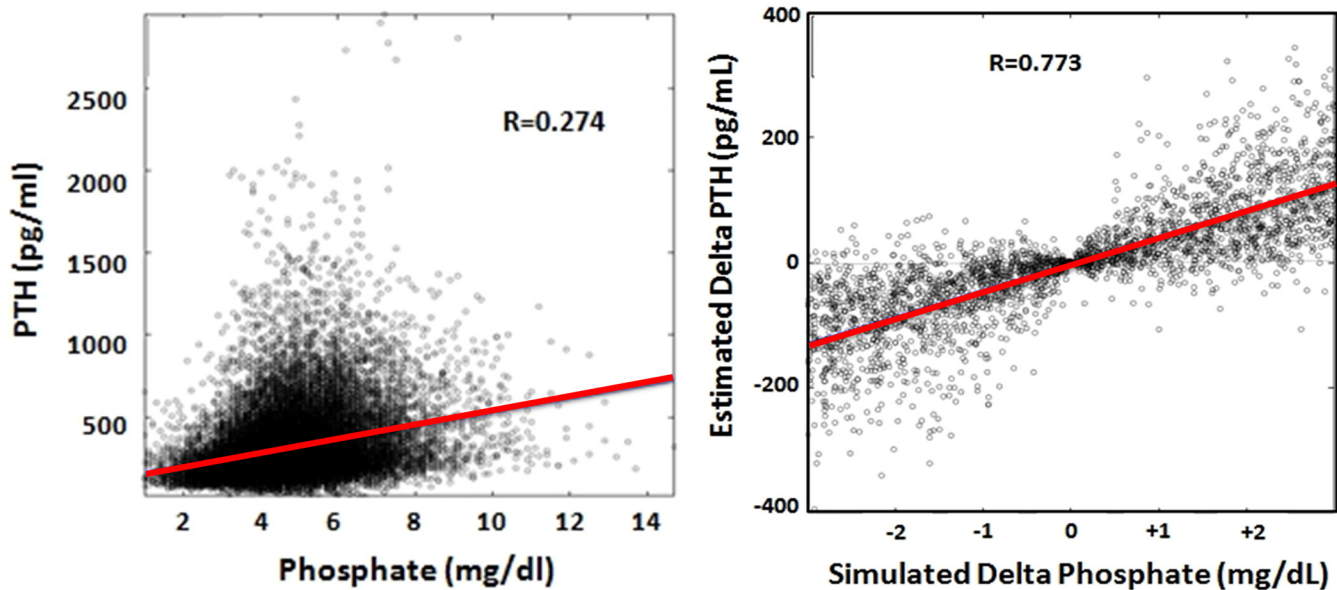


Fig 5 (A)

Fig 5 (B)

Fig 5. (A) Correlation of Serum PTH and phosphate concentration. (B) Estimated change in PTH concentration in response to a simulated change in serum phosphate concentration. These estimations are derived from Random Forest regression models for the prediction of outcome variables.

doi:10.1371/journal.pone.0146801.g005

patients. The parameters evaluated were the serum concentrations of Ca, P and PTH, parameters that guide the physician in deciding therapy in daily practice [1,2,21,22,23,24,25]. One of the basis of the present study is the assumption that these three parameters should not be evaluated individually because the processes involved in their regulation are common, thus changes in one parameter will affect the others. One should be open to accept that mathematical models, other than classical linear statistical analysis, may be helpful to analyze variables that are closely related to each other.

While mutual influences between serum levels of Ca, P and PTH are expected, it is difficult to quantify to what extent these parameters may influence each other because of the interference of confounding factors. RF analysis can identify non-linear relationships among many variables without any a priori assumption, i.e. the probability distribution that characterizes

Table 4. Comparison of prediction error among variables included in the feature analysis for phosphate.

	PTH	AP	G	Age	Vint.	Paric	Ca Bind.	Sev
Ca	<0.001	<0.001	ns	<0.001	<0.001	ns	Ns	ns
PTH	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
AP	-	-	<0.001	<0.001	ns	<0.001	<0.001	<0.001
G	-	-	-	<0.0001	<0.001	ns	Ns	ns
Age	-	-	-	-	<0.001	<0.001	<0.001	<0.001
Vint.	-	-	-	-	-	<0.001	<0.001	<0.001
Paric	-	-	-	-	-	-	Ns	ns
Ca Bind.	-	-	-	-	-	-	-	ns

Abbreviations: (PTH) Parathyroid Hormone (P) Phosphate; (AP) Alkaline Phosphatase, (G) Gender, (Vint) Vintage, (CTR) Calcitriol, (Paric) Paricalcitol, (Ca Bind.) Calcium binders, (Sev.) Sevelamer. (CM) Calcimimetics, (Ca) Calcium

doi:10.1371/journal.pone.0146801.t004

the phenomenon is automatically estimated by the data itself, while with a classical analytical approach the knowledge of data distribution has to be assumed *a priori* [16,17]. Therefore, taking into account the complexity and non-linearity involved in the CKD-MBD scenario the use of an advanced RF model resulted in a significant improvement in extracting useful information from the available data. In particular through the derived predictive models of Ca, PTH and P it was possible to identify interesting patterns of associations among different parameters and therapies. Even if a specific association is not straightforward, the predictive model is able to mediate the concurrent effect of all involved features making the association to emerge clearly. Predictive models may eventually help clinical diagnosis and eventually therapeutic decisions. Evidently final decisions must include bedside assessment of patient condition and we are far away from obtaining valid algorithms that could replace an experienced Doctor. Further studies will be necessary to prove that these advanced methods of analysis have substantial value in assessing complex relationships of various parameters in a single patient. If that is the case, this type of information may help to decide therapeutic interventions. The Ca, PTH and P models obtained by RF do not expose any previously unknown association, however these models allows to quantify to what extent one independent variable is needed to optimally predict values of the outcome variable.

The study reveals that serum calcium is positively associated with time on dialysis but the association becomes negative after 8 years on dialysis. This finding is explained by the fact that serum calcium is negatively associated with age; it is likely that after several years of dialysis, patients are older and the serum calcium decreases. Therefore this finding is explained by the negative age effect overcome the positive effect of time in hemodialysis. What is really relevant is that removing age and vintage from the model that predict serum Ca will cause an error of 0.57 and 0.58 mg/dl respectively (Fig 2 and Table 2). The Serum Ca concentration is directly related to both PTH and alkaline phosphatase suggesting that in most patients PTH drives the serum Ca concentration. Without PTH or alkaline phosphatase the error in predicting Ca would be more than 0.55 mg/dl. Ca is also associated to the use of calcimimetics and vitamin D analogs. [26,27,28].

This study found that PTH was positively associated with dialysis vintage (Fig 3 and Table 3); this is not an unusual finding since there are reports showing that hyperparathyroidism progresses with time on dialysis until the patient becomes older. Age, is known to be negatively associated with PTH and with a tendency to adynamic bone disease [10,29].

The close association between PTH and alkaline phosphatase illustrates the consistency of the RF technique used in the present work. It was interesting to find out that the PTH-P association was superior to the PTH-Ca association.

The positive association between serum PTH concentration and serum P levels is shown in Fig 5(A) and 5(B). Plotting the absolute values of serum PTH and P concentration (Fig 5A) shows a correlation that although significant is not valid to predict changes in PTH concentration based on changes in serum P level. The prediction of a change in PTH resulting from a modification of serum P is show in Fig 5(B), this correlation is obtained using RF approach. This figure shows the correlation between PTH and P being analyzed as the deviation of PTH keeping fixed all the features and varying just P. The results of this analysis reveal a strong correlation between the modifications in PTH and P values. A modification in one of these two parameters should be accompanied by a corresponding variation in the other parameter [2,18,30]. High serum P could be associated to high PTH through different mechanisms: a high P level may directly increase PTH secretion and synthesis, high phosphate also produces skeletal resistance to PTH [31,32]. Additionally high PTH level increases bone efflux of P that contributes to hyperphosphatemia. This strong association between PTH and P suggest that strategies to decrease PTH and P should be combined.

High PTH levels were associated to the use of calcimimetics, this association is expected since calcimimetics are used to reduce PTH levels [26,27]. Calcitriol and paricalcitol were also associated to values of PTH, however it was not as strong as the PTH-calcimimetics association.

Values of serum P concentration were negatively associated with age. This has been reported previously [33] and it has been attributed to a decrease in protein intake. Interestingly, it appears, from the present analysis, that age may influence serum concentration of PTH, P and also Ca but guidelines do not mention any type of correction for age. It may be appropriate to investigate whether there is an age level at which these parameters are being significantly modified. The majority of patients were on one or more types of phosphate binders and, according to the performed analyses, it appears that the type of binder is not linked to a given phosphate level.

In conclusion the analysis of complex interactions between the different parameters of mineral metabolism may benefit from more advanced mathematical approach such as Random Forest which is able to identify and computes non-linear relationships among many variables without any *a priori* assumption.

Supporting Information

S1 Dataset. Compilation of data.
(ZIP)

Acknowledgments

We are grateful to Francesco Bellocchio for the time and effort analyzing the dataset.

Author Contributions

Conceived and designed the experiments: MR MDS AMM PA. Performed the experiments: MR MDS AMM CB. Analyzed the data: MR MDS AMM CB FM RIM PC PA. Contributed reagents/materials/analysis tools: CB FM RIM MDS. Wrote the paper: MR MDS AMM PA.

References

1. KDIGO clinical practice guideline for the diagnosis, evaluation, prevention, and treatment of Chronic Kidney Disease—Mineral and Bone Disorder (CKD-MBD). *Kidney Int* 2009; 113(suppl):S1–S130.
2. Ketteler M, Elder JG, Evenepoel P, Ix JH, Jamal SA, Lafage-Proust MH, et al. Revisiting KDIGO clinical practice guideline on chronic kidney disease—mineral and bone disorder: a commentary from a Kidney Disease: Improving Global Outcomes controversies conference. *KID Int*. 2015.
3. Block GA, Kilpatrick RD, Lowe KA, Wang W, Danese MD. CKD-mineral and bone disorder and risk of death and cardiovascular hospitalization in patients on hemodialysis. *Clin J Am Soc Nephrol*. 2013; 8:2132–40. doi: [10.2215/CJN.04260413](https://doi.org/10.2215/CJN.04260413) PMID: [24052218](https://pubmed.ncbi.nlm.nih.gov/24052218/)
4. Young EW, Albert JM, Satayathum S, Goodkin DA, Pisoni RL, Akiba T, et al. Predictors and consequences of altered mineral metabolism: the Dialysis Outcomes and Practice Patterns Study. *Kidney Int* 2005; 67:1179–1187. PMID: [15698460](https://pubmed.ncbi.nlm.nih.gov/15698460/)
5. Block GA, Klassen PS, Lazarus JM, Ofsthun N, Lowrie EG, Chertow GM. Mineral metabolism, mortality, and morbidity in maintenance hemodialysis. *J Am Soc Nephrol* 2004; 15:2208–2218. PMID: [15284307](https://pubmed.ncbi.nlm.nih.gov/15284307/)
6. Tentori F, Fuller DS, Port FK, Bieber BA, Robinson BM, Pisoni RL. The DOPPS practice monitor for US dialysis care: potential impact of recent guidelines and regulatory changes on management of mineral and bone disorder among US hemodialysis patients. *Am J Kidney Dis*. 2014; 63:851–4. doi: [10.1053/j.ajkd.2014.02.003](https://doi.org/10.1053/j.ajkd.2014.02.003) PMID: [24613057](https://pubmed.ncbi.nlm.nih.gov/24613057/)
7. Floege J, Kim J, Ireland E, Chazot C, Druke T, de Francisco A, et al. ARO Investigators. Serum iPTH, calcium and phosphate, and the risk of mortality in a European haemodialysis population. *Nephrol Dial Transplant* 2011; 26:1948–1955. doi: [10.1093/ndt/gfq219](https://doi.org/10.1093/ndt/gfq219) PMID: [20466670](https://pubmed.ncbi.nlm.nih.gov/20466670/)

8. Vervloet MG, Massy ZA, Brandenburg VM, Mazzaferro S, Cozzolino M, Ureña-Torres P, et al. On behalf of the CKD-MBD Working Group of ERA-EDTA. Bone: a new endocrine organ at the heart of chronic kidney disease and mineral and bone disorders. *Lancet Diabetes Endocrinol*. 2014; 2:427–436. doi: [10.1016/S2213-8587\(14\)70059-2](https://doi.org/10.1016/S2213-8587(14)70059-2) PMID: [24795256](https://pubmed.ncbi.nlm.nih.gov/24795256/)
9. Evenepoel P, Rodriguez M, Ketteler M. Laboratory Abnormalities in CKD-MBD: Markers, Predictors, or Mediators of Disease? *Semin Nephrol*. 2014; 34:151–163. doi: [10.1016/j.semnephrol.2014.02.007](https://doi.org/10.1016/j.semnephrol.2014.02.007) PMID: [24780470](https://pubmed.ncbi.nlm.nih.gov/24780470/)
10. Palomares I, Ramos R, Martin-Malo A, Merello JI, Praga M, Luño J, et al. ORD Group. Compliance with mineral metabolism targets in hemodialysis patients: moving backwards? *Blood Purif*. 2013; 36:122–31. doi: [10.1159/000353418](https://doi.org/10.1159/000353418) PMID: [24217176](https://pubmed.ncbi.nlm.nih.gov/24217176/)
11. Kalantar-Zadeh K, Kuwae N, Regidor DL, Kovesdy CP, Kilpatrick RD, Shinaberger CS, et al. Survival predictability of time-varying indicators of bone disease in maintenance hemodialysis patients. *Kidney Int* 2006; 70:771–780. PMID: [16820797](https://pubmed.ncbi.nlm.nih.gov/16820797/)
12. Danese MD, Belozeroff V, Smirnakis K, Rothman KJ. Consistent control of mineral and bone disorder in incident hemodialysis patients. *Clin J Am Soc Nephrol*. 2008; 3:1423–9. doi: [10.2215/CJN.01060308](https://doi.org/10.2215/CJN.01060308) PMID: [18596117](https://pubmed.ncbi.nlm.nih.gov/18596117/)
13. Jamal SA, Vandermeer B, Raggi P, Mendelsohn DC, Chatterley T, Dorgan M, et al. Effect of calcium-based versus non-calcium-based phosphate binders on mortality in patients with chronic kidney disease: an updated systematic review and meta-analysis. *Lancet*. 2013; 382:1268–77. doi: [10.1016/S0140-6736\(13\)60897-1](https://doi.org/10.1016/S0140-6736(13)60897-1) PMID: [23870817](https://pubmed.ncbi.nlm.nih.gov/23870817/)
14. Cleophas TJ, Zwinderman AH. *Machine Learning in Medicine*. Springer, 2013.
15. Savage N. Better medicine through machine learning. *Communications of the ACM*, vol. 55, no. 1, pp. 1719, 2012.
16. Breiman L. *Random Forests*. *Machine Learning*, vol. 45, pp. 5–32, 2001.
17. Genuer R, Poggi JM, Tuleau-Malot C. Variable selection using random forests. *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2225–2236, 2010.
18. Leung RK, Wang Y, Ma RC, Luk AO, Lam V, Ng M, et al. Using a multi-staged strategy based on to predict genotype-phenotype risk patterns in diabetic kidney disease: a prospective case-control cohort analysis. *BMC Nephrol*. 2013 Jul 23; 14:162. doi: [10.1186/1471-2369-14-162](https://doi.org/10.1186/1471-2369-14-162) PMID: [23879411](https://pubmed.ncbi.nlm.nih.gov/23879411/)
19. Scian MJ, Maluf DG, Archer KJ, Turner SD, Suh JL, David KG, et al. Identification of biomarkers to assess organ quality and predict posttransplantation outcomes. *Transplantation*. 2012 Oct 27; 94(8):851–8. doi: [10.1097/TP.0b013e318263702b](https://doi.org/10.1097/TP.0b013e318263702b) PMID: [22992769](https://pubmed.ncbi.nlm.nih.gov/22992769/)
20. Mas VR, Mas LA, Archer KJ, Yanek K, King AL, Gibney EM, et al. Evaluation of gene panel mRNAs in urine samples of kidney transplant recipients as a non-invasive tool of graft function. *Mol Med*. 2007 May-Jun; 13(5–6):315–24. PMID: [17622313](https://pubmed.ncbi.nlm.nih.gov/17622313/)
21. Torregrosa JV, Bover J, Cannata Andía J, Lorenzo V, de Francisco AL, Martínez I, et al. Spanish Society of Nephrology recommendations for controlling mineral and bone disorder in chronic kidney disease patients (SEN-MBD). *Nefrologia* 2011; 31(suppl 1):3–32. doi: [10.3265/Nefrologia.pre2011.Jan.10816](https://doi.org/10.3265/Nefrologia.pre2011.Jan.10816) PMID: [21468161](https://pubmed.ncbi.nlm.nih.gov/21468161/)
22. K/DOQI clinical practice guidelines for bone metabolism and disease in chronic kidney disease. *Am J Kidney Dis* 2003; 42(suppl 3): S1–S201.
23. Al Aly Z, González EA, Martin KJ, Gellens ME. Achieving K/DOQI laboratory target values for bone and mineral metabolism: an uphill battle. *Am J Nephrol* 2004; 24: 422–426. PMID: [15308874](https://pubmed.ncbi.nlm.nih.gov/15308874/)
24. Cozzolino M, Ureña-Torres P, Vervloet MG, Brandenburg V, Bover J, Goldsmith D, et al. Is Chronic Kidney Disease-Mineral Bone disorder (CKD-MBD) Really a Syndrome? *Nephrol Dial Transplant*. 2014; 29(10): 1815–1820. doi: [10.1093/ndt/gft514](https://doi.org/10.1093/ndt/gft514) PMID: [24516228](https://pubmed.ncbi.nlm.nih.gov/24516228/)
25. Bhan I, Dubey A, Wolf M. Diagnosis and management of mineral metabolism in CKD. *Journal of General Internal Medicine*. 2010 Jul; 25(7):710–716. doi: [10.1007/s11606-010-1316-y](https://doi.org/10.1007/s11606-010-1316-y) PMID: [20352364](https://pubmed.ncbi.nlm.nih.gov/20352364/)
26. Block GA, Martin KJ, de Francisco AL, Turner SA, Avram MM, Suranyi MG, et al. Cinacalcet for secondary hyperparathyroidism in patients receiving hemodialysis. *N Engl J Med*. 2004; 350:1516–25. PMID: [15071126](https://pubmed.ncbi.nlm.nih.gov/15071126/)
27. Chertow GM, Block GA, Correa-Rotter R, Drüeke TB, Floege J, Goodman WG, et al. EVOLVE Trial Investigators. Effect of cinacalcet on cardiovascular disease in patients undergoing dialysis. *N Engl J Med*. 2012; 367:2482–94. doi: [10.1056/NEJMoa1205624](https://doi.org/10.1056/NEJMoa1205624) PMID: [23121374](https://pubmed.ncbi.nlm.nih.gov/23121374/)
28. Palmer SC, McGregor DO, Craig JC, Elder G, Macaskill P, Strippoli GF. Vitamin D compounds for people with chronic kidney disease not requiring dialysis. *Cochrane Database Syst Rev*. 2009 Oct 7.
29. Lertdumrongluk P, Lau WL, Park J, Rhee CM, Kovesdy CP, Kalantar-Zadeh K. Impact of age on survival predictability of bone turnover markers in hemodialysis patients. *Nephrol Dial Transplant*. 2013; 28:2535–45. doi: [10.1093/ndt/gft290](https://doi.org/10.1093/ndt/gft290) PMID: [23904397](https://pubmed.ncbi.nlm.nih.gov/23904397/)

30. Felsenfeld AJ, Rodriguez M. Phosphorus, regulation of plasma calcium, and secondary hyperparathyroidism: a hypothesis to integrate a historical and modern perspective. *J Am Soc Nephrol.* 1999; 10:878–90. PMID: [10203374](#)
31. Almaden Y, Hernandez A, Torregrosa V, Canalejo A, Sabate L, Fernandez Cruz L, et al. High phosphate level directly stimulates parathyroid hormone secretion and synthesis by human parathyroid tissue in vitro. *J Am Soc Nephrol.* 1998; 9:1845–52. PMID: [9773785](#)
32. Berdud I, Martín-Malo A, Almaden Y, Tallon S, Concepcion MT, Torres A, et al. Abnormal calcaemic response to PTH in the uraemic rat without secondary hyperparathyroidism. *Nephrol Dial Transplant.* 1996; 11:1292–8. PMID: [8672025](#)
33. Lorenzo V, Martín M, Rufino M, Jiménez A, Malo AM, Sanchez E, et al. Protein intake, control of serum phosphorus, and relatively low levels of parathyroid hormone in elderly hemodialysis patients. *Am J Kidney Dis.* 2001; 37:1260–6. PMID: [11382697](#)