Multilinear Regression Analysis of Sweat Secretion Volumes in Cystic Fibrosis Patients

Paola Lecca
University of Trento, Italy
paola.lecca@unitn.it

Marina Bertini, Roberta Vicentini, Emily Pintani, Ciro D'Orazio
Cystic Fibrosis Centre,
Azienda Ospedaliera Universitaria Integrata Verona, Italy

Valeria Esposito, Karina Kleinfelder, Claudio Sorio
University of Verona, Italy
{valeria.esposito, karina.kleinfelderfontanesi}@univr.it
{claudio.sorio}@univr.it

Paola Melotti
Cystic Fibrosis Centre,
Azienda Ospedaliera Universitaria Integrata Verona, Italy,
paola.melotti@aovr.veneto.it

***Abstract*—Personalized medicine approach in cystic fibrosis (CF) is focusing on detection of cystic fibrosis transmembrane conductance regulator (CFTR) function in single patients and standardized outcomes for CFTR function in vivo. We applied Optical Sweat Rate Beta Adrenergic (OSRBA) test for measuring sweat rates in individual human sweat glands. The results were analyzed according to a multilinear regression model in non-CF, healthy carriers (HTZ), CF patients; two groups of these were tested during treatment with CFTR modulators such as lumacaftor/ivacaftor (Orkambi) or PTC 124 (Ataluren). We found that different sets of statistically significant coefficients of the multilinear regression of the volume of sweat secretory glands characterize different CFTR genotype as well as different responses to different pharmacological treatments.**

I. Introduction

Cystic Fibrosis (CF) is a chronic, genetic disease that affects the function and abundance of secretory glands, called the cystic fibrosis transmembrane conductance regulator (CFTR), that are located in the apical membrane of epithelial cells. These membrane proteins are responsible for the production of fluids like mucus, tears, saliva, sweat and digestive enzymes. The CF development is related to mutations in the CFTR gene that is inherited as an autosomal recessive disorder, in which a person inherits one defective CFTR allele from each parent. When there is only one defective allele, the person is a CF carrier, but when there are two mutated CFTR alleles, the patient develops CF [1], [2], [3], [4], [5]. Due to abnormal function of the secretory glands, patients experience an excessive production of mucus that affects the respiratory, digestive and reproductive systems. This abnormal production of sticky and thick mucus is particularly dangerous in the lungs since the mucus accumulates and block the airways, causing persistent lung infections and limiting the ability to breathe over time [6], [7], [8]. The incidence of CF varies across the globe, the United States are among the countries with higher incidence of CF. The disease is particularly common also among Caucasians of Northern European descent and among Latinos and American Indians, especially the Pueblo and Zuni. The incidence of CF in the European Union is around 1 in 2000-3000 newborns [1], [9], [10]. On the contrary, the disease seems to be severely under-diagnosed in Asia, so that the statistics for the cases of CF in this country may depict the realistic scenario. The golden standard diagnostic test for CF is the Gibson and Cooke method [11]. A test for concentration of electrolytes in sweat in cystic fibrosis of the pancreas utilizing pilocarpine by iontophoresis for measuring sweat chloride concentration ($>$60mmol/L in CF; $<$ 40mmol/L in non CF) [12]. Subjects with values in the borderline range (40-60mmol/L) require other CFTR function tests with higher linearity. ORSBA has been developed as complementary for supporting controversial diagnosis as well as for detecting improvements of CFTR function during treatments with CFTR modulators [13], [14], [15], [16]. Multicenter validation of in vivo CFTR assays supporting controversial diagnosis and detection of CFTR function improvement during CFTR targeted therapy require a reproducible procedure as well as reliable readouts. Along this line, the recently introduced CFTR-targeted drugs open to a personalized therapy approach in CF and require standardized outcomes for CFTR function *in vivo*. In this study we propose a multilinear model to describe the behaviour of the average ratio between the volume of secretory sweat droplets induced by CFTR dependent stimuli and the volume of those formed by CFTR independent sweating (response), as function of a set of variables including the number of glands, the volume sweat droplets from single glands, the concentration of chloride in sweat, and the expiratory volume of the patient. We present the results of the regression diagnostics and used the model to select the independent variables that best describe the response.

II. The experiments

Sweat secretion rates were given by changes of volume (nl/min) of sweat droplets secreted on the forearm in an oil layer, including the presence of a water-soluble blue dye (erioglaucine disodium crystals). We computed a ratio between CFTR-dependent, evoked by intradermal microinjection of a β-adrenergic cocktail (Cktl), and CFTR-independent, induced by methacoline as cholinergic stimulus (MCh), sweat secretion rates by multiple individual glands. We tested a number N of patients as in Table I.

A multilinear regression model of mean Cktl/MCh ratio (dependent variable) vs gender, number of Cktl/MCh activated sweat droplets, number of MCh activated droplets, Cktl/MCh glands, sweat chloride measured by Gibson and Cooke method [11], lung function (FEV1) has been implemented and tested.

ISSN 2305-7254

TABLE I. PHARMACOLOGICAL TREATMENTS AND NUMBER OF TESTED PATIENTS.

Condition/Treatment	Meaning	N
Kalydeco	ivacaftor (CFTR modulator)	2
Orkambi	ivacaftor+lumacaftor (CFTR modulator)	38
PTC124	Ataluren, PTC124 (CFTR modulator)	17
CRD	CFTR Related Disorder	9
UNK	controversial diagnosis	8
DCP	primary ciliary dyskinesia	5
BPCO	Bronchial Pulmonary Chronic Obstructive disease	4
CF	Cystic fibrosis	59
CTR	Control	32
HTZ	heterozygous	33
PTC T0 + Orkambi T0	before treatment with Ataluren or Orkambi	27

The mean CM ratio is the average ratio between the volume of sweat droplets formed in two phases, i.e phase Cktl and phase MCh (hereafter called Mean CM_ratio). The experimental test to monitor the secretory function of CFTR consists of two sequential periods of stimulated secretion (the test monitors and compares the sweat produced by secretion that is CFTR-dependent with the non-dependent secretion of CFTR). The first period (10 min) measures sweating MCh (at the response to MCh; CFTR-independent) and the second period (30 min) measures sweating Cktl (at the cocktail response; CFTR-dependent). The increase in volume of each individual identified gland was measured over time in patients with different CFTR genotype and in different experimental conditions determined by different pharmacological treatments. The dependent variable is the outcome of the ORSBA test [17].

We used the model to identify the set of explicative predictors in all the groups reported in the first columns of Table I. The dependent variable is the outcome of the ORSBA test [17]. In the next section we report a comprehensive description of the data used to build the model.

III. THE DATA AND MULTILINEAR MODEL

The multiple linear regression equation is as follows:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \qquad (1)$$

where $\hat{Y}$ is the predicted or expected value of the dependent variable (hereafter called *response*), X_1 through X_n are n distinct independent or predictor variables, β_0 is the value of $\hat{Y}$ when all of the independent variables (X_1 through X_n) are equal to zero, and β_1 through $\beta_n X$ are the estimated regression coefficients. ϵ, the error term, is an independent random variables with a normal distribution of mean 0 and variance σ^2 [18], [19], [20].

Each regression coefficient represents the change in $\hat{Y}$relative to a one unit change in the respective independent variable. In the multiple regression situation, β_1, for example, is the change in $\hat{Y}$ relative to a one unit change in X_1, holding all other independent variables constant (i.e., when the remaining independent variables are held at the same value or are fixed). Statistical tests can be performed to assess whether each regression coefficient is significantly different from zero. The predictors X_j that are multiplied by regression coefficients not significantly different from zero (p-value greater than the significance threshold) are discarded, because they are judged variables whose contribution to the variability of the response is not significant [21].

In our study, the response is the Mean CM_ratio and its candidate predictors are:

- **Mch_glands**: number of glands induced to produce sweat droplets at 10 min after methacholine (Mch)-sweating stimulation via intradermal injection (M phase).
- **Cktl_glands**: number of glands induced to secrete sweat droplets at 30 min. after cocktail (Cktl)-sweating stimulation via intradermal injection (C phase).
- **CM_ratio_glands**: ratio between the total number of glands induced in Cktl phase with the total number of sweat droplets formed in MCh phase.
- **Mch_rate** (nl/min): average of volume of sweat droplets formed in MCh phase.
- **Cktl_rate** (nl/min): average of volume of sweat droplets formed in Cktl phase.
- **FEV1** (forced expiratory volume in the 1st second): the volume of air that can be forced out in one second after taking a deep breath.
- **Sweat_Cl** (mmol/L): concentration of chloride that is excreted in sweat.
- **Gender** of the patient.

We partitioned the initial data frames into three data frames according to the number of mutations of CFTR. We thus obtained three sub-data frames DF_1, DF_2 and DF_3 with the following sample size: 0 mutations (non-CF) - 77 data points; 1 mutation (healthy carriers) 27 data points; and, 2 mutations (CF) 228 data points. Each data frame DF_i $(i = 1, 2, 3)$ has been further partitioned into sub-data frames $p^{(i)}$ $SDF_j^{(i)}$ $(j = 1, 2, \cdots, p^{(i)})$ according to the pharmacological treatments. In our analysis the sample size of a data frame corresponds to the number of performed tests and not to the number of patients (indicated in table I). The minimum sample size required by this analysis is 10. Moreover, a putative predictor is not included in the linear model analysis if the number of non-available data ("NA") for it is greater that the 50% of the sample size.

The computational pipeline of the multilinear regression analysis consisted of the following steps:

1) fit a multilinear model inclusive of all putative predictors to the experimental data
2) test the significance of individual regression coefficients β_i by performing a two-sided t-test to test the null hypothesis $H_0 : \beta_i = 0$ against the alternative hypothesis $H_1 : \beta_i \neq 0$; a decision is taken according to the p-values of the test on each single regression coefficient
3) detection of multicollinearity through the calculation of variance inflation index (VIF); the threshold is traditionally set to 5 [22]
4) check for nonlinearities by inspecting the partial residual plots
5) estimation of residuals autocorrelation; indeed the residuals are assumed to be uncorrelated with one

another, which implies that the Y's are also uncorrelated. If residuals are significantly autocorrelated, the mean square error may be seriously underestimated. The consequence of this is that the standard errors are underestimated, the t-tests show significance when there is none, and the confidence intervals are smaller than they should be. Therefore any hypothesis tests or confidence intervals that require the use of the t or F distribution are invalid.
The assumption of absence of residuals' autocorrelation indicates the specification of a linear model could be not appropriate to the data.

The mathematical foundations and the algorithmic procedure implementing steps 1), 2) and 5) are very well known to the majority of the practitioners also without a mathematical education. The steps 3) and 4) instead use less known analysis tools, although they are of particular importance in multilinear regression diagnostics. They are especially useful in our case study where we use a multilinear model and its diagnostics as (i) explorative tools for the identification of the best predictors, and as (ii) indicators of relationships other than linear. In the rest of this section, we give a definition of the VIF index and of the partial residual plots.

Variance Inflation Factor to detect multicollinearity
Multicollinearity means very high correlation among the predictors. If present in the data, multicollinearity jeopardizes the reliability of the statistical inferences made about the data. Multicollinearity is generally caused by the inclusion in a model of a variable which is computed/correlated from/to other variables in the data set, and it can result in several problems as follows.

- The partial regression coefficient may not be estimated precisely, i.e their standard errors are likely to be high. Moreover, the confidence intervals of the regression coefficients tend to become very large and the t statistics tend to be very small. Therefore it becomes unlikely to reject the null hypothesis H_0.
- Multicollinearity may manifest itself in a change in the signs and magnitudes of the partial regression coefficients from one sample to another sample.
- In presence of multicollinearity, the outcome of the t statistic on the single regression coefficient is not significant but the overall outcome of the statistic (e.g. the F statistics) is significant.
- As a consequence, multicollinearity makes it very hard to assess the relative importance of the candidate predictors in explaining the variation of the response.

Multicollinearity can be detected by estimating the variance inflation factor.

The variances of the estimated coefficients are inflated when multicollinearity exists. So, the variance inflation factor for the estimated coefficient β_k is just the factor by which the variance is inflated [22]. For simplicity, suppose that X_k is the only predictor, then

$$\hat{Y}_i = \beta_0 + \beta_k X_{ik} + \epsilon_i. \tag{2}$$

where $i = 1, \dots n$ ranges over the number of predictors. It can be shown that the variance of the estimated coefficient β_k is:

$$\text{Var}_{\text{min}}(\beta_k) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_{ik} - \overline{x}_k)}. \tag{3}$$

Suppose to have a model with correlated predictors:

$$\hat{Y}_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \dots \beta_{n-1} X_{i,(n-1)} + \epsilon_i.$$

If some of the predictors are correlated with the predictor X_k, then the variance of β_k is inflated, and becomes:

$$\text{Var}(\beta_k) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_{ik} - \overline{x}_k)} \times \frac{1}{1 - R_k^2} \tag{4}$$

where R_k^2 is the R^2-value obtained by regressing the k-*th* predictor on the remaining predictors. The greater the linear dependence among the predictor X_k and the other predictors, the larger the R_k^2 value, the larger the variance of β_k. The Variance Inflation Factor quantify how much larger the variance of the coefficient β_k becomes in case of multicollinearity [22] and thus

$$\text{VIF}(\beta_k) = \frac{\text{Var}(\beta_k)}{\text{Var}_{\text{min}}(\beta_k)} = \frac{1}{1 - R_k^2} \tag{5}$$

A VIF of 1 means that there is no correlation among the k-*th* predictor and the remaining predictor variables, and hence the variance of bk is not inflated at all. The general rule of thumb is that VIFs exceeding 4 suggest further investigations, while VIFs exceeding 10 indicate serious multicollinearity requiring model correction [22].

Partial Residual Plots
When performing a linear regression with a single independent variable, a scatter plot of the response variable against the independent variable provides a good indication of the nature of the relationship. Usually, in order to have an indication of the nature of the relationship, scatter plots of the response variable against the independent variables are provided. However, in a single scatter plot the effect of the other independent variables in the model is not taken into account. Partial residual plots (often called Component plus Residuals plots) instead show the relationship between a given independent variable and the response variable given that other independent variables are also in the model [23], [24], [25], [26]. On a partial residual plot, we plot

$$\text{Residuals} + \beta_i X_i \quad \text{versus} \quad X_i. \tag{6}$$

where Residuals are the residuals from the full model. The partial residual plot allows to show the relationship between a given explanatory variable and the response variable by excluding the influence of the other predictors of the model. Consequently, this kind of plot is capable to identify the predictors linked to the response by a nonlinear function, and suggests the relevant refinements of the model. For instance, suppose that the k-*th* predictor results non-linearly correlated to the response, then, if f is the best candidate function describing this nonlinearity, the model has to be amended as

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + f(X_k) + \cdots \beta_n X_n + \epsilon. \tag{7}$$

IV. RESULTS

We report the results of the regression and regression diagnostics for the three principal subsets of the input data frame, i.e for the data relative to 0, 1, and 2 CFTR mutations (respectively, CTR, HTZ, and CF groups), and for the pharmacological treatments to which CF patients have been subjected.

We checked whether the assumptions of linear regression hold true. By plots (not reported in this paper) visualizing the residual errors we checked: (i) non-linearity of the response - predictor relationships (residual vs fitted plots), (ii) normal distribution of the residuals (Q-Q plot); (iii) heteroscedasticity, i.e. non-constant variance of error terms (scale-location plot); (iv) presence of influential values in the data that can be: outliers: extreme values in the response variable, and high-leverage points, i.e extreme values in the predictors variable (residuals vs leverage plot). We found that all these assumptions are satisfied for the data of CTR, and HTZ group, but not for the CF, Orkambi and PTC 124 groups, in which presence of non-linearities in the response and heteroscedasticy are brought to light and indicate the need of a more accurate analysis to identify which predictors are the main responsible of the non-linear behaviour of the response. The absence of significant residuals autocorrelation (i.e no model misspecification) is testified by the results of the Durbin-Watson test [27] in Table II.

TABLE II. TWO-SIDED DURBIN-WATSON (D-W) TEST [27] TO DETECT RESIDUALS AUTOCORRELATION. THE NULL HYPOTHESIS IS THAT THE AUTOCORRELATION $\rho = 0$.

Group	ρ	D-W statistics	p-value
CTR	0.0717	1.854	0.47
HZT	-0.0354	2.070	0.756
CF	-0.0714	2.140	0.758
Orkambi	0.0115	1.966	0.612
PTC 124	-0.3831	2.625	0.208

The main results of our analysis are shown in Figs 1-5. We indicate on barplots the values of the p-values of the t statistics on each single regression coefficients (plot A.), useful to identify the significant predictors, the values of the VIF (plot B.), useful to detect collinear predictors, and the partial residual plots (plots C.) useful to identify the predictors mainly responsible for non-linearity of the response. The list of characteristics of the predictors of Mean_CM_ration for each group is reported in Tables III and IV.

TABLE III. THE RESPONSE MEAN_CM_RATIO IN DIFFERENT GROUPS IS DESCRIBED BY DIFFERENT SETS OF CANDIDATE UNCORRELATED PREDICTORS.

Group	Independent predictors VIF < 5
CTR	Gender, MCh_rate, Cktl_rate
HTZ	Gender, MCh_rate, Cktl_rate
CF	Gender, MCh_rate, Cktl_rate
Orkambi	Gender, MCh_glands, MCh_rate, Cktl_rate, FEV1, Sweat_Cl
PTC 124	Gender, MCh_glands, MCh_rate, FEV1, Sweat_Cl

We found that while in the control group (CTR) and in the group of heterozygotes (HTZ) the sets of best predictors share MCh_Rate and Cktl_rate, but not Gender. Interestingly, Gender is a good predictor only in HTZ group (p-value less tha 10% and VIF < 5.). Gender is reliably not significant in CTR group. The reliability of the estimate of its significance is ensured by the fact that multilinear model is a good mode for the patients of this group. The estimate of the significance of the Gender variable in the other groups may not be accurate since in these groups non-linear dependences are revealed by this analysis. Consequently, further analysis, and a refinement of the model specifying the appropriate non-linear dependencies of the response on the best predictors are necessary to achieve an accurate estimate of the p-value of the t statistics.

TABLE IV. THE RESPONSE MEAN_CM_RATIO IN DIFFERENT GROUPS IS DESCRIBED BY DIFFERENT SETS OF SIGNIFICANT PREDICTORS. ONLY FOR THE CONTROL GROUP ALL THE PREDICTORS ARE LINEARLY CORRELATED TO THE RESPONSE.

Group	Best predictors p-value $< 10\%$	Nonlinearities partial residual plots
CTR	MCh_rate, Cktl_rate	–
HTZ	MCh_rate, Cktl_rate, Gender	MCh_glands
CF	Cktl_glands, CM_ratio_glands, Cktl_rate	Cktl_rate
Orkambi	Cktl_glands	MCh_glands, Cktl_rate, CM_ratio_glands
PTC 124	–	all the predictors

In the group of patients affected by CF, the set of the best predictors is instead strongly different from the predictors' set of CTR and HTZ gropus. Cktl_rate is a good predictor in CTR, HTZ, and CF groups, but while in CTR and HTZ patients it correlates linearly with the response, in the group of CF patients it exhibits a non linear behaviour (see Table IV and Fig. 3 plot C.). Finally, we found that the multilinear model does not properly describe the response. The two pharmacological treatments (Orakambi and PTC 124) induces non-linear behaviours of the response as function of most of the candidate predictors. The determination of the uniqueness and the analytical expression of the nonlinear function of the predictors that explains the progress of the response is not a problem of easy solution with the data currently available. The results warn us about strong possible deviations from the linearity that can be expressed and validated with the collection of samples of greater numbers, but above all they can be read as co-predictors of genotypes that lead to the development of the disease and or pharmacological interventions.

All these results are consistent with the expected effects of drugs targeting CFTR mutations in CF patients and the phenotypes of the different groups of healthy carriers, CF patients and non-CF subjects.

V. CONCLUSIONS

This paper presents the results of a joint effort between medical doctors, biologists, and statisticians for the collection and analysis of data concerning the sweat secretion volumes in patients affected by cystic fibrosis, healthy carriers and controls. An advanced experimental technology combined with a multi-linear regression and diagnostic performed with modern techniques has allowed a first collection and understanding of the data. Furthermore, it revealed the existence of possible mathematical relationships more complex than linear ones in cases of disease and pharmacological treatments. The agreement with the expectations of the effect of the treatments and the different genotypes has made us confident in the goodness of the analysis tools developed so far.

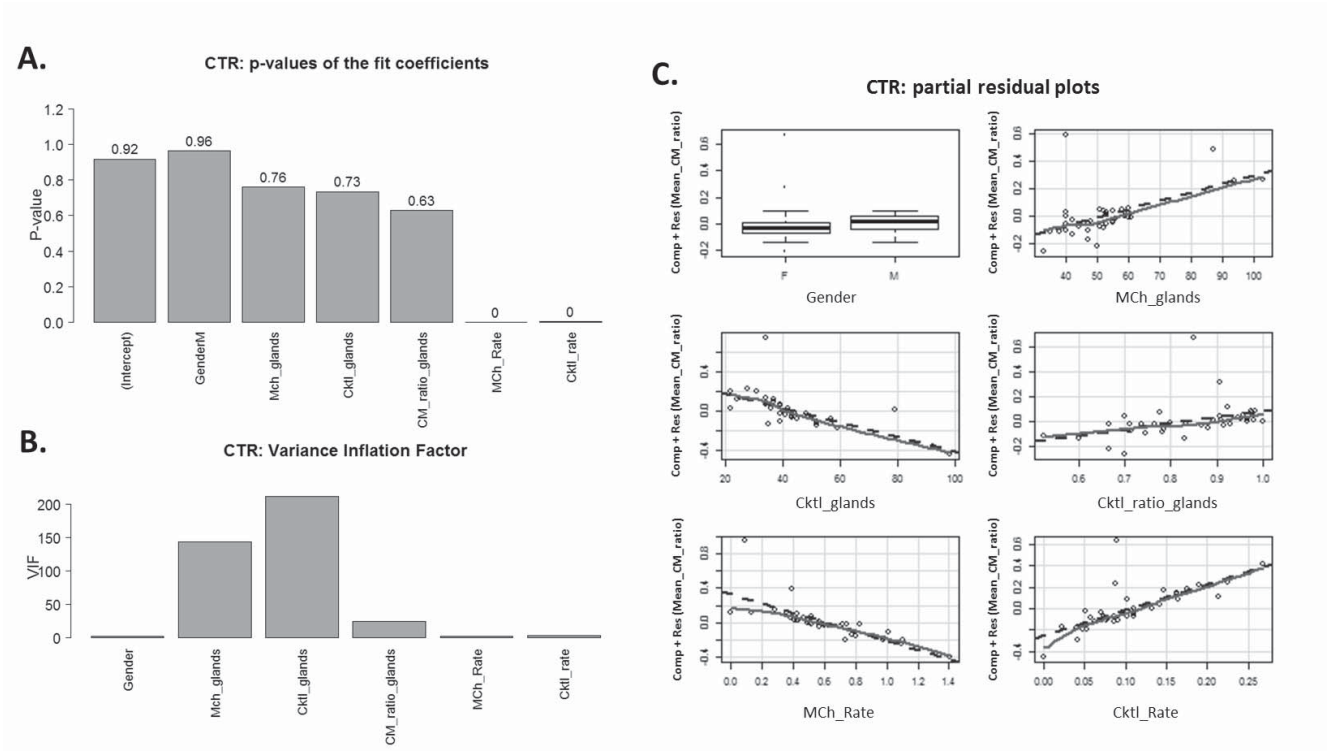

Fig. 1. For the group of control (CTR) patients, the best predictors according to the p-values are MCh_rate, Cktl_rate (plot A.); the predictor that do not exhibit multicollinearity are Gender, MCh_rate, Cktl_rate (plot B); and, all the predictors are linearly correlated to the response (plot C.).

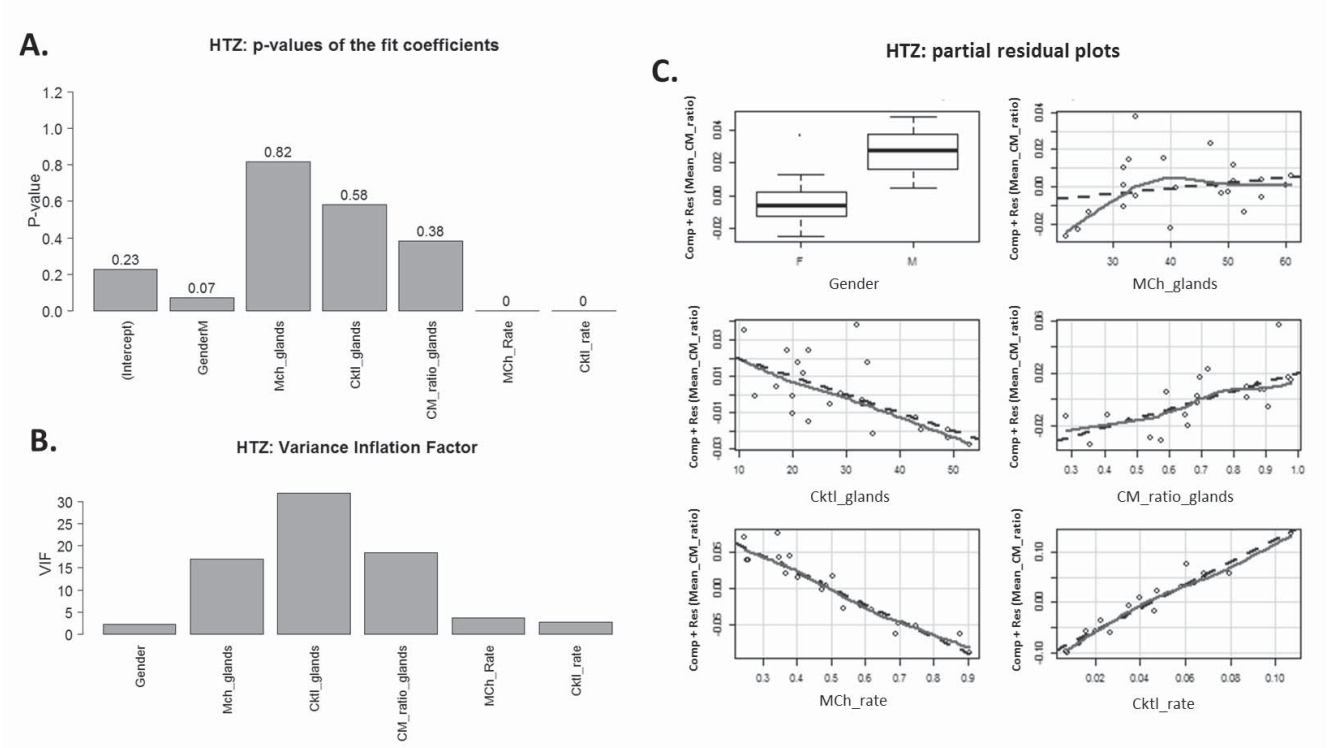

Fig. 2. For the group of healthy carriers (HTZ), the best predictors according to the p-values are MCh_rate, and Cktl_rate (plot A.); the predictor that do not exhibit multicollinearity are Gender, MCh_rate, Cktl_rate (plot B.); and, all the predictors, except that MCh_glands, are linearly correlated to the response (plot C.).

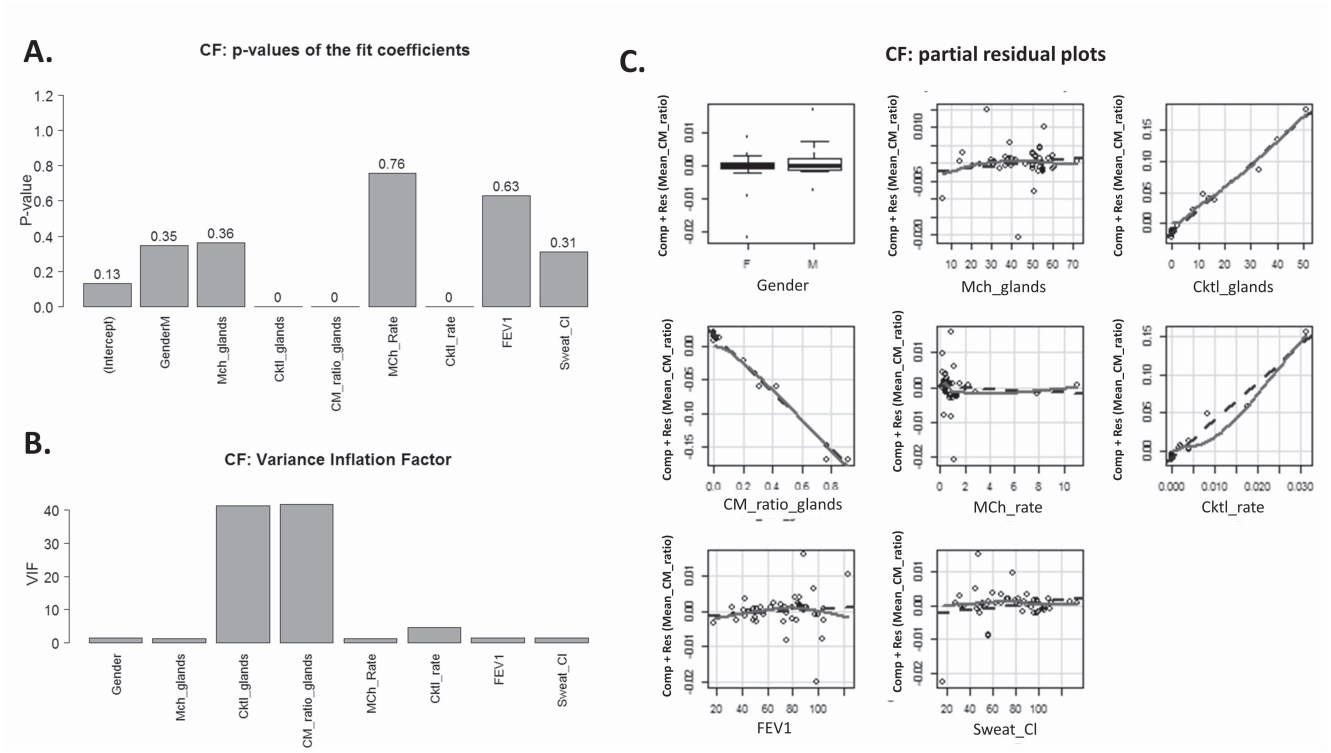

Fig. 3. For the group of CF patients, the best predictors according to the p-values are Cktl_glands, CM_ratio_glands, Cktl_rate (plot A.); the predictor that do not exhibit multicollinearity are Gender, MCh_glands, MCh_rate, Cktl_rate, FEV1, Sweat_Cl (plot B.); and, all the predictors, except that Cktl_rate, are linearly correlated to the response (plot C.).

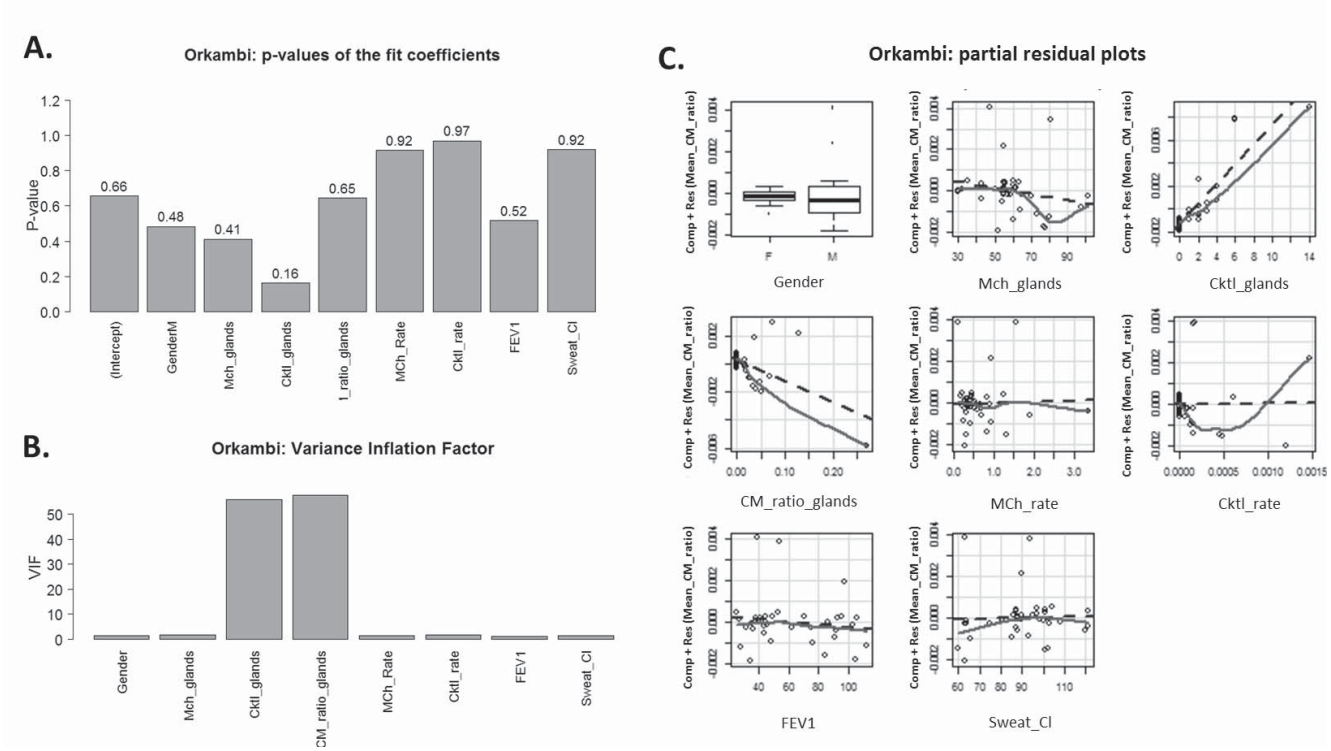

Fig. 4. For the CF patients treated with Orkambi, the best predictor according to the p-values is (approximately) Cktl_glands with a p-value of 0.16 (plot A.); the predictor that do not exhibit multicollinearity are Gender, MCh_glands, MCh_rate, Cktl_rate, FEV1, Sweat_Cl (plot B.); and, the predictors non-linearly correlated to the response are MCh_glands, Cktl_rate, CM_ratio_glands, and Sweat_Cl (plot C.).

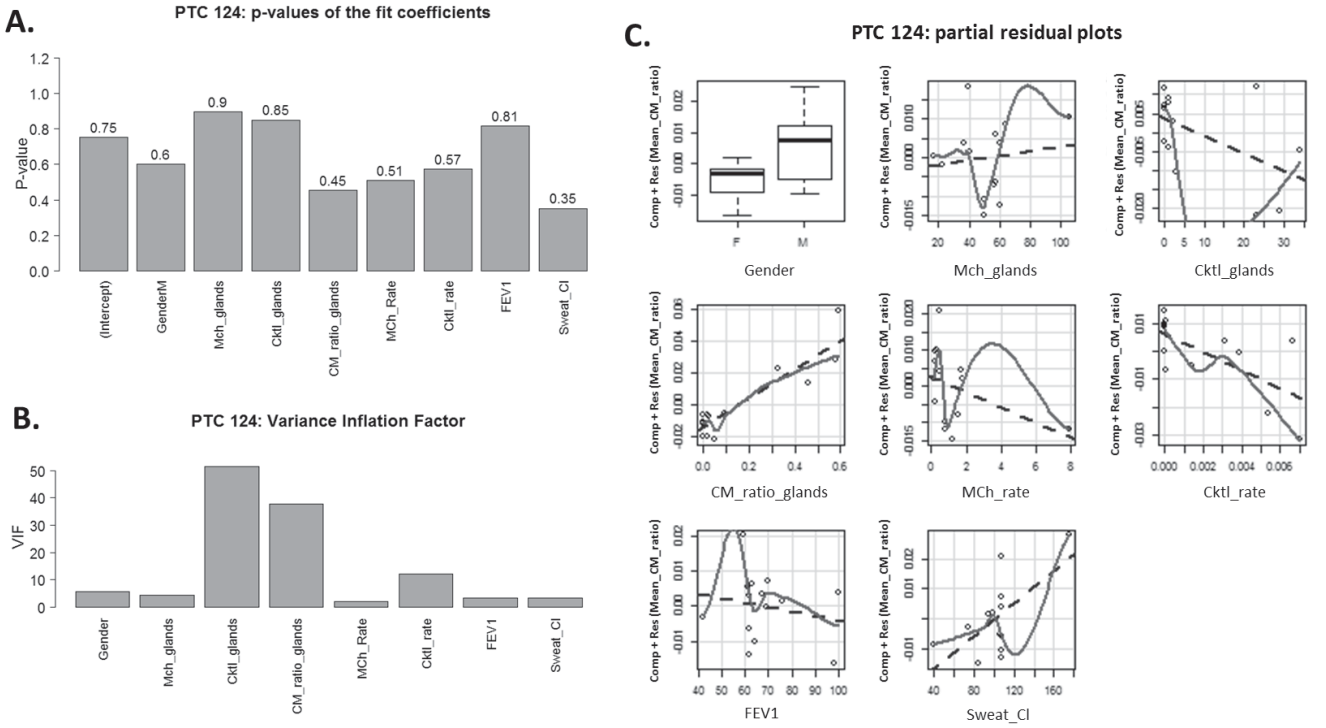

Fig. 5. For the CF patients treated with PTC 124, the multilinear model for the Mean_CM_ration is not adequate. Non of the candidate predictors are significant (plot A.), the majority of predictors are not collinear (plot B.), but they are not linearly correlated to the response (plot C.).

ACKNOWLEDGEMENTS

The authors thank all volunteers tested in this study. The study was supported by Lega Italiana Fibrosi Cistica - Associazione Veneta Onlus and Italian Cystic Fibrosis. Research Foundation grant#5/2016: Delegazioni di Imola e Romagna.

REFERENCES

[1] V. Saint-Criq and M. A. Gray, Role of CFTR in Epithelial Physiology, Cell. Mol. Life Sci. 74:93115, 2017.

[2] S. L. Martiniano, S. D. Sagel, and E. T. Zemanick, Cystic fibrosis: a model system for precision medicine, Curr Opin Pediatr. 28(3): 312317, 2016

[3] M. L. C. Anselmo, Cystic fibrosis: Overview, In: Taussig LML, LI., editor. Pediatric Respiratory Medicine. 2nd. Elsevier Health Sciences, p. 845-857, 2008

[4] M. J. Welsh, Bonnie W. Ramsey, F. Accurso, G. R. Cutting, Cystic fibrosis. In: Scriver CRB, AL.; Sly, WS.; Valle, D., editors. The Metabolic and Molecular Bases of Inherited Disease. 8th. McGraw-Hill; p. 5121-5188, 2001.

[5] G. R. Cutting, Cystic fibrosis genetics: From Molecular Understanding to Clinical Application. Nat Rev Genet. 16:4556, 2015.

[6] H. Matsui, B. R. Grubb, R, Tarran, S. H. Randell, J. T. Gatzy, C. W. Davis, R. C. Boucher, Evidence for Periciliary Liquid Layer Depletion, not Abnormal Ion Composition, in the Pathogenesis of Cystic Fibrosis Airways Disease, Cell 95: 10051015, 1998.

[7] I. J. Haq, M. A. Gray, J. P. Garnett, C. Ward, M. Brodlie, Airway Surface Liquid Homeostasis in Cystic Fibrosis: Pathophysiology and Therapeutic Targets, Thorax 71: 284287, 2016.

[8] A. A. Pezzulo, X. X. Tang, M. J. Hoegger, M. H. Abou Alaiwa, S. Ramachandran, T. O. Moninger, P. H. Karp, C. L: Wohlford-Lenane, H. P. Haagsman, M. van Eijk, B. Banfi, A. R. Horswill, D. A. Stoltz, P. B. Jr McCray, M. J. Welsh, J. Zabner, Reduced Airway Surface pH Impairs Bacterial Killing in the Porcine Cystic Fibrosis Lung, Nature 487: 109113, 2012.

[9] G. M. Solomon, S. G. Marshall, B. W. Ramsey, S. M. Rowe, Breakthrough therapies: Cystic Fibrosis (CF) Potentiators and Correctors, Pediatr Pulmonol 50, Suppl 40: S3S13, 2015.

[10] J. S. Elborn, Cystic fibrosis, Lancet 388: 25192531, 2016.

[11] L. Gibson, R. Cooke, A Test for Concentration of Electrolytes in Sweat in Cystic Fibrosis of the Pancreas Utilizing Pilocarpine by Iontophoresis, Pediatrics, 23(3):545-9, 1959.

[12] P. Lebecque, T. Leal, C. De Boeck, M. Jasper, H. Cuppens, J. J. Cassiman, Mutation of the Cystic Fibrosis Gene and Intermediate Sweat Chloride Levels in Children, Am J Respir Crit Care Med;165:757-61, 2002.

[13] J. E. Char, M. H. Wolfe, H. J. Cho, I. H. Park, J. H. Jeong, E. Frisbee, C. Dunn, Z. Davies, C. Milla, R. B. Moss, E. A. C. Thomas, J. J. Wine, A Little CFTR Goes a Long Way: CFTR-Dependent Sweat Secretion From G551D and R117H-5T Cystic Fibrosis Subjects Taking Ivacaftor, Plos ONE 9(2):e88564, 2014.

[14] J. J. Wine, J. E. Char, J. Chen, H. J. Cho, C. Dunn, E. Frisbee, Nam Soo Joo, Carlos Milla, Sara E. Modlin, Il-Ho Park, Ewart A. C. Thomas, Kim V. Tran, Rohan Verma, Marlene H. Wolfe, In Vivo Readout of CFTR Function: Ratiometric Measurement of CFTR-Dependent Secretion by Individual, Identifiable Human Sweat Glands, Plos ONE 8(10):e77114, 2013.

[15] J. K. Behm, G. Hagiwara, N. J. Lewiston, P. M. Quinton, J. J. Wine, Hyposecretion of Beta-adrenergically Induced Sweating in Cystic Fibrosis Heterozygotes, Pediatric research 22(3): 271-6, 1987.

[16] P. Quinton, L. Molyneux, W. Ip, A. Dupuis, J. Avolio, E. Tullis E, D. Conrad, A. K. Shamsuddin, P. Durie, T. Gonska, Beta-adrenergic Sweat Secretion as a Diagnostic test for Cystic Fibrosis, American Journal of Respiratory and Critical Care Medicine 186(8): 732-9, 2012.

[17] G. Bergamini, G. Tridello, E. Calcaterra, S. Ceri, M Tagliasacchi, F. Bianchi, F. Monti, A. Masciadri, E. Laudanna, D. Peserico, E. Sorio, V. Esposito, T. Leal, B. M. Assael, C. Sorio, P. Melotti, J Cyst Fibros. 17(2):186-189, 2018.

[18] J. F. Hair, R. Anderson, R. L. Tatham, W.C. Black, "Multivariate Data Analysis", Upper Saddle River, NJ: Prentice Hall, 2006.

[19] M. H. Kutner, C. J. Nachtsheim, J. Neter, "Applied Linear Regression Models" (4th ed.) McGraw-Hill Irwin, 2004.

[20] G. D. Garson, Multiple Regression, Statistical Publishing Associates, 2014.

[21] L. L. Nathans, F. L. Oswald, K. Nimon, K. "Interpreting Multiple Linear Regression: A Guidebook of Variable Importance", Practical Assessment, Research & Evaluation, 17, 123-136, 2015.

[22] M. Akinwande, H. Dikko, H. A. Samson, "Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis", Open Journal of Statistics, 5, 754-767, 2015.

[23] T. Ryan, "Modern Regression Methods", John Wiley, 1997.

[24] J. Neter, W. Wasserman, and M. H. Kunter, "Applied Linear Statistical Models", 3rd ed., Irwin, 1990.

[25] N. R. Draper and H. Smith, "Applied Regression Analysis", 3rd. ed., John Wiley, 1998.

[26] Joseph W. McKean and Simon J. Sheather, "Exploring Data Sets Using Partial Residual Plots Based on Robust Fits" Lecture Notes-Monograph Series Vol. 31, L_1-Statistical Procedures and Related Topics (1997), pp. 241-256.

[27] J. Fox, Applied Regression Analysis and Generalized Linear Models, Third Edition, Sage, 2016.