C. Costa[1], P. Menesatti[1°], M.A. Brighetti[2], A. Travaglini[2], V. Rimatori[1], A. Di Rienzo Businco[3], S. Pelosi[4], A. Bianchi[5], P.M. Matricardi[5,6], S. Tripodi[3]

Pilot study on the short-term prediction of symptoms in children with hay fever monitored with e-Health technology

[1]Consiglio per la Ricerca e la sperimentazione in Agricoltura, Unità di ricerca per l'ingegneria agraria, Rome, Italy
[2]Department of Biology, University of Rome "Tor Vergata", Rome, Italy
[3]Department of Pediatrics and Unit of Pediatric Allergology, Sandro Pertini Hospital, Rome, Italy
[4]TPS Production, Rome, Italy
[5]Department of Pediatrics, Mazzoni Hospital, Ascoli Piceno, Italy
[6]Department of Pediatric Pneumology and Immunology, Charité Medical University, Berlin, Germany

Corresponding author

Salvatore Tripodi
Department of Pediatrics and Unit of
Pediatric Allergology
Sandro Pertini Hospital,
via dei Monti Tiburtini 385
00157 Rome, Italy
E-mail: salvatore.tripodi@gmail.com

Summary

Forecasting symptoms of pollen-related allergic rhinoconjunctivitis at the level of individual patients would be useful to improve disease control and plan pharmacological intervention. Information Technology nowadays facilitates a more efficient and easier monitoring of patients with chronic diseases. We aimed this study at testing the efficiency of a model to short-term forecast symptoms of pollen-AR at the "individual" patient level. We analysed the data prospectively acquired from a group of 21 Italian children affected by pollen-related allergic rhinoconjunctivitis and recorded their symptoms and medication "Average Combined Score" (ACS) on a daily basis during April-June 2010-2011 through an informatics platform (Allergymonitor™). The dataset used for prediction included 15 variables in four categories: (A) date, (B) meteo-climatic, (C) atmospheric concentration of 5 pollen taxa, and (D) intensity of the patient's IgE sensitization. A Partial Least Squares Discriminant Analysis approach was used in order to predict ACS values above a fixed threshold value (0.5). The best performing predicting model correctly classified 77.8% ± 10.3% and 75.5% ± 13.2% of the recorded days in the model and test years, respectively. In this model, 9/21 patients showed ≥ 80% correct classification of the recorded days in both years. A better performance was associated with a higher degree of patient's atopic sensitization and a time lag > 1. Symptom forecasts of seasonal allergic rhinitis are possible in highly polysensitised patients in areas with complex pollen exposure. However, only predictive models tailored to the individual patient's allergic susceptibility are accurate enough. Multicenter studies in large population samples adopting the same acquisition data system on smart phones are now needed to confirm this encouraging outcome.

Conflict of interest

Simone Pelosi is CEO of TPS Production; Salvatore Tripodi is a shareholder in TPS Production; Paolo M. Matricardi is receiving lecture fees from Allergopharma, Thermo Fisher Scientific and TPS Production. The rest of the authors declare that they have no relevant conflicts of interest.

Introduction

Millions of people worldwide, particularly children, suffer from allergic rhino-conjunctivitis (AR) induced by pollens (pollen-AR) (1,2). AR negatively affects patients' performance of daily activities, sleep patterns, cognitive function, work and school productivity and quality of life. Less than half of the pa-

tients regularly follow medical advice, drug therapies generally achieve only partial control of symptoms, and patient's adherence to therapy is often poor (3).

The symptoms of pollen-AR appear when the concentration of the offending pollen reaches a "threshold" value so that avoiding or reducing exposure to pollens would be useful. Awareness about airborne concentrations of pollens helps the patients and their doctors to plan effective prevention and treatment and to improve adherence to drug therapy (4). Consequently, forecasting symptoms of pollen-AR at individual level would also be useful to improve disease control (5). Unfortunately, threshold values vary not only among patients but also in the same patient during the pollen season. In fact, symptoms severity is dependent not only on pollen exposure, but also on patient's specific factors, such as living environment, the level of IgE antibodies, sensitization and simultaneous exposure to other allergenic sources, and the clinical reactivity of the target organ (eyes, nose, lungs) (6,7). Patients and doctors should be helped understanding how symptoms change during a pollen season; this may help identifying the individual co-factors facilitating symptomatic manifestations and, consequently, disease self-management. Until now, many scales, indexes or scores have been created to measure the severity of AR and the impact of this disease on the patient's daily life. The most frequently used are symptom scores (SS), medication scores (MS), and combined symptom-medication scores (SMS) (8,9).

Information Technology nowadays facilitates a more efficient and easier patient monitoring (7). Applications have been used to forecast symptoms at patient group (clustering based on pollen concentrations and allergic symptoms) (7) and - most importantly - individual level (10). The design and the development of patient-specific prediction models is a challenging task (7). The relationship between pollen counts and measures of disease severity can be simple and linear (11,12), but also very complex and non-linear (12,13). Then adequate mathematical tools (e.g. forecasts models) are necessary. Algorithms and complex models are being increasingly applied to predict trends of chronic diseases; they are rapidly evolving and their complexity is increasing with the number of variables taken into account (14,15,16). Thanks to this evolution, the performance of forecast models continues to improve in many research fields (14,15,16,17,18). Forecasting models of pollen allergies that incorporate information about the individual patient's susceptibility are moving their first steps with encouraging results (7) and - as for other chronic disease - there is room for improvement of their performance.

We aimed this pilot study at testing the efficiency of a model to short-term forecast symptoms of pollen-AR at the "individual" patient level. This model is based not only on meteo-climatic data and pollen concentrations, but also on individual risk-fac-

tors (hence the name), such as the patient's molecular profile and the overall intensity of IgE sensitization. Eventually, we analysed the data prospectively acquired from a group of children affected by hay fever and using on a daily basis and for two consecutive seasons an informatics platform (Allergymonitor™) to monitor allergic symptoms according to internationally established criteria.

Materials and methods

Study population and study design

The study population consisted of patients seeking care for pollen-AR at the Pediatric Allergy Outpatient Unit of the Sandro Pertini Hospital in Rome. Inclusion criteria were the following: A) a diagnosis of pollen-AR; B) IgE sensitization to one or more of the following four pollen sources: birch, grass, olive, pellitory, i.e. the most relevant ones in Rome between April and June (19); C) the intention to stay in Rome for the whole study period; D) lack of sensitization to perennial allergens such as *Dermatophagoides pteronyssinus*, cat, dog, *Alternaria alternata* or other molds. Each patient underwent skin prick tests for *Dermathophagoides pteronyssinus*, *Phleum pratense*, *Cynodon dactylon*, pellitory, mugwort, ragweed, cypress, birch, plane, *Olea europaea*, cat, and dog, with Histamine 0,1 mg/ml and glycerol solution as positive and negative control respectively (ALK-Abellò Milan, Italy), and a blood sample was drawn to test the concentration of IgE to major pollen allergenic molecules.

After parents or legal tutors gave a written informed consent, the patients were asked to record daily on a web-platform (AllergyMonitor©, TPS production, Rome, Italy) their symptoms and medication during the pollen season (from April 1st April to June 30th) both in 2010 and in 2011. Only patients recording symptoms and medications for > 20 consecutive days during the examination period were examined. No interpolation was applied to missing data, and only consecutive days were considered. This study was embedded in a larger epidemiological study on pollen allergy that was approved by the Ethic Committee of the Sandro Pertini Hospital (20).

Definitions

Pollen-AR was diagnosed in the presence of: (1) nasal and/or eye symptoms (apart from common cold) (21) for at least three weeks during one of the two last pollen seasons, and (2) positive SPT (wheal reaction > 3 mm) in accordance with clinical history and local pollination period. Pollen-AR was classified as mild or moderate/severe, as well as intermittent or persistent (ARIA classification) (22). The age at onset of pollen-AR was reported by the parents as their child's age as the first year with relevant symptoms. The duration of AR since its onset was established

C. Costa, P. Menesatti, M.A. Brighetti, A. Travaglini, V. Rimatori, A. Di Rienzo Businco, S. Pelosi, A. Bianchi, P.M.

218   Matricardi, S. Tripodi

as the difference in years between the child's age at recruitment and the child's age at AR onset. Asthma was classified as intermittent, mild persistent, moderate persistent or severe persistent (GINA classification) (23).

Symptom-Medication-Score

Patients were asked to record their symptoms and medication once a day (at evening) on Allergymonitor™. This tool is a user-friendly web based platform to monitor allergic rhinitis and asthma, accessed both via mobile (iOS and Android operative systems) or PC. For the purposes of the present study, the "Average Combined Score" (ACS) (9) has been automatically and daily calculated by the platform. We chose this score because it is derived by a combination of symptoms' score and drug therapy, so it is more reliable. The patients recorded their ocular, nasal, and bronchial symptoms as well as medication. The ACS index was calculated as previously reported (9) by combining the Rhinoconjunctivitis Total Symptom Score (RTSS) and the Rescue Medication Score (RMS) according to the following formula: [(RTSS / 6 + RMS) / 2]. The RTSS includes six individual symptoms: four nasal (sneezing, rhinorrhoea, itching, and congestion) and two ocular (itching and tearing). The intensity of each symptom can be expressed with a value from 0 to 3: 0 = absent (no sign/symptom evident); 1 = mild (sign/symptom clearly present, but minimal awareness; easily tolerated); 2 = moderate (definite awareness of sign/symptom that is bothersome, but tolerable); 3 = severe (sign/symptom that is hard to tolerate; causes interference with activities of daily living and/or sleeping). The daily total of six symptoms combined can reach a score between 0 and 18. The RMS adopts a scale based on the type of drug taken: 0 = no drug, 1 = antihistamines (topical and/or oral), 2 = nasal corticosteroids, 3 = oral corticosteroids. If two or more types of drugs are taken in a given day, the one with the highest score is taken into account. The resulting total score ranges from 0 to 3, but the score is a continuous one and its level of precision is normally at the level of 2 decimals and it is as sensitive as the other scores (24). A patient was considered to have symptoms in a given day if he/she had an ACS of at least 0.5 in that day. Allergymonitor© calculates the ACS, extracts data and automatically generates databases usable for outsource analyses.

Parameters and databases

The whole dataset (x-block) was composed of 15 variables, including (A) the ordinal dates of the year, (B) five meteo-climatic variables [temperature (maximum, minimum and mean; °C), wind speed (mean; Km h⁻¹) and rain rate (mm)], (C) the air concentration of the five most representative pollens of the study area [Betulaceae (*Betula* blooming period is May-June), Cory-

laceae (*Carpinus* and *Ostrya*, April-May), Graminaceae (April-June), Oleaceae (*Olea*, May-June) and Urticaceae (blooming throughout the year)], and (D) four scores based on the intensity of the patient's IgE sensitization, expressed as specific activity (see below) to the major allergenic protein of four pollens multiplied for the daily counts of the corresponding pollen (IgE-pc). These 15 variables were then used in the modeling approaches (see below). The 15 variables were used to compose three datasets progressively including an increasing number of variables (**table 1**): (1) the simplest or "meteo" dataset, composed only by ordinal dates and the 5 meteo-climatic variables (six parameters; DMC), (2) the intermediate or "meteo-pollen" dataset, including the meteo parameters and the concentration of the five pollens taken into account (overall 11 parameters; DMCP), (3) the global or "meteo-pollen-IgE" dataset, including also the four pollen-sensitization indexes (overall 15 variables; DMCPI).

Table 1 - *Database of increasing complexity used to predict trends of symptoms in 21 patients with hay fever.*

Variables	DMC	DMCP	DMCPI
Date (ordina¹)	X	X	X
Meteoclimatic1	X	X	X
Pollen concentrations²		X	X
IgE-pc³			X

1 including five variables: temperature (maximum, minimum and mean), wind speed (mean; Km h-1) and rain rate (mm)
2 including five variables: Betulaceae, Corylaceae, Graminaceae, Oleaceae and Urticaceae
3 including four variables: index of sensitization to Corylaceae, Graminaceae, Oleaceae and Urticaceae

Meteorological data and pollen counts

Both meteorological data and pollen counts were recorded at the meteorological and aerobiological station of the University of Rome "Tor Vergata", located at a distance of 10.5 km from the study center, with validated methodologies (25,26). The monitoring station aerobiological pertains to the Italian Network of Monitoring Aerobiology, R.I.M.A.®, coordinated by the Italian Association of Aerobiology® (AIA®). A volumetric sampler type Hirst (27) Model 2000 VPPS Lanzoni (28) has been used. The data acquisition is routinely carried out according to standard procedures (Standard UNI 11008:2004 - "Qualità dell'aria - Metodo di campionamento e conteggio dei granuli pollinici e delle spore fungine aerodisperse") and the pollen counts are reported as daily concentration and expressed in grains/m³ air (26).
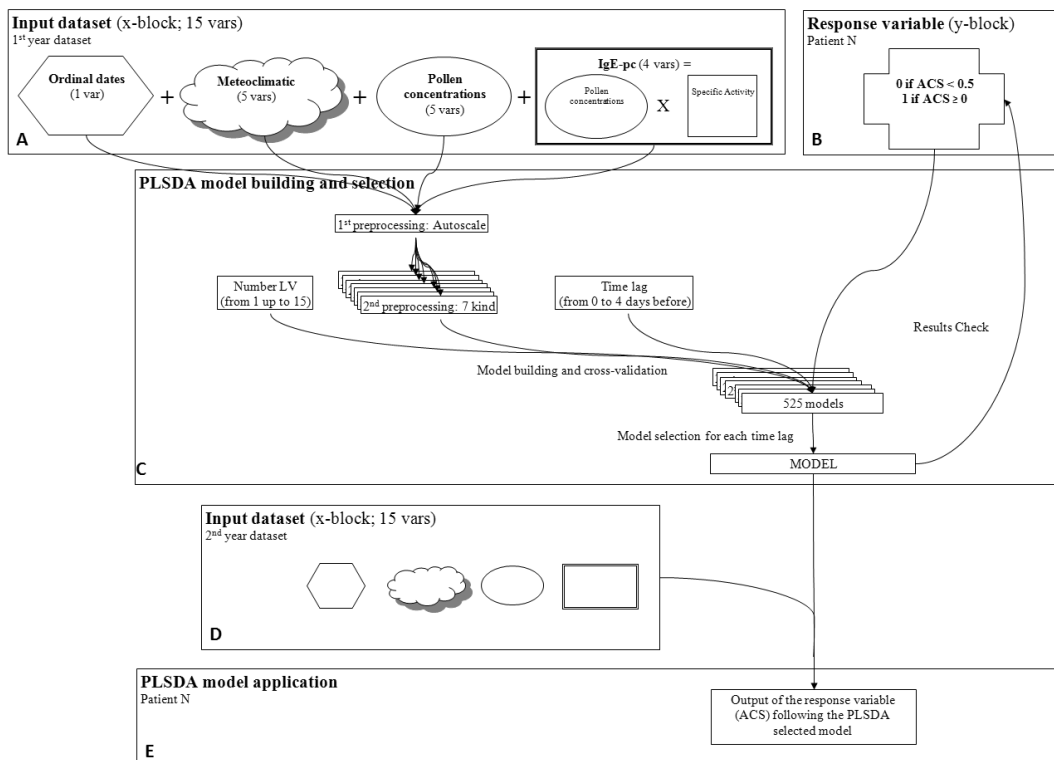
IgE assays

Total and specific IgE determination for this study have been performed as previously reported (29). IgE for allergenic molecules were tested in sera of patients showing a wheal reaction > 2 mm elicited by the corresponding allergenic source by ImmunoCAP FEIA (ThermoFisher Scientific, Uppsala, Sweden). The following major allergenic molecules were selected as previously suggested: Graminaceae (*Phleum pratense*, rPhl p1 and rPhl p5b), Oleaceae (*Olea europaea*, nOle e 1), Betulaceae (*Betula verrucosa*, rBet v1), Urticaceae (*Parietaria judaica*, rPar j2). Results were expressed in kU/L. Detection ranged from 0.35 kU/l to 100 kU/l. The IgE specific activity (SA) is the fraction (%) of patient's serum concentration of specific IgE antibodies to a given allergenic molecule within the total IgE immunoglobulins (sIgE/tIgE). For example, the SA of IgE to rPhl p1 in a patient with a level of 30 kU/l of serum IgE to rPhl p1 and a level of 360 kU/l of serum total IgE is 8,33% (30*100/360). Specific activity of IgE antibodies is a good marker for predicting the clinical response to specific allergen-specific immunotherapy (30).
An index obtained by multiplying the pollen concentration with the specific activity was also created and defined "IgE-pc" (Pollen concentration multiplied by the Specific Activity). For each patient, the value of IgE against rBet v1, rPhlp1 and rPhlp 5b, nOle e1, and rPar j2 has been multiplied by the value of the daily pollen concentration of the corresponding pollen type (Betulaceae and Corylaceae for rBet v1, Graminaceae for rPhl p1 + rPhl p5b, Olea for nOle e1, Urticaceae for rPar j2).

The multivariate time lag modeling

The proposed multivariate modeling approach predicts up to 4 days before the event the presence or the absence of symptoms (ACS > 0.5) (y-block) from an input dataset (x-block). A flowchart of the multivariate time lag modelling approach on the DMCPI dataset is summarized in **figure 1**. A Partial Least Squares Discriminant Analysis (PLSDA) approach was used in order to predict ACS values above a fixed threshold value (0.5). PLSDA consists of a classical Partial Least Squares (PLS) regression analysis where the response variable is categorical (Y-block), 0 if ACS < 0.5 (considered as absence of symptoms) or 1 if ACS ≥ 0.5 (presence of symptoms), thus expressing the class membership of the statistical units (31,32). The partitioning design consists, for each patient, in using the data from the year with the

Figure 1 - *Flowchart of the structure of the input/output dataset and the multivariate time lag modelling approach.*

C. Costa, P. Menesatti, M.A. Brighetti, A. Travaglini, V. Rimatori, A. Di Rienzo Businco, S. Pelosi, A. Bianchi, P.M.
Matricardi, S. Tripodi

220

larger number of records to build and cross-validate the dataset (hereafter labeled as the "model year"), and the other year as independent test (hereafter labeled as the "test year"). The prediction ability of PLSDA also depends on the number of latent vectors (LV) used in the model. The x-block was preprocessed using an autoscale algorithm (centres columns to zero mean and scales to unit variance). For the model development, a second row preprocessing step was applied. At least 7 different kind of row (second) preprocessing were applied: none, baseline (Weighted Least Squares), detrend (remove a linear trend), mean centering, msc (multiplicative scatter correction with offset), normalize (normalization of the rows) and snv (Standard Normal Deviate).

For each dataset, the best models were extracted at 5 different forecast levels (time lag) ranging from 0 to 4 days. The time lag, represents the gap between the ACS (y-block) and the x-block shifted i days before (16). For example at time lag = 2, using the today x-block variables the ACS relative to 2nd following day was forecasted. Therefore, for best model selection the following modeling parameters were considered: time lag (from 0 to 4), number of LVs (from 1 up to 15) and x-block second pre-processing (7). This leads to a total of 4,410 (210 for each patient) potential models for dataset DMC, 8,085 (385 for each patient) potential models for dataset DMCP and 11,025 (525 for each patient) potential models for dataset DMCPI to be elaborated. The different number of models depends on the maximum number of LV (equal to the number of the variables) that could be used. As selection rule, for each patient the 5 models (one for each time lag) with the mean higher performance value (percentage of correct classification for both model and test sets) were considered. For DMCPI model sensitivity and specificity parameters were calculated. The models were developed using a procedure written in the MATLAB 7.1 R14 environment. Difference among means has been tested using the t-test, $p < 0.05$ was considered significant. Normality has been tested with the Shapiro-Wilk test. Confidence intervals have been expressed as standard deviation (SD) or standard error (SE).

Observation period and missing values

The period used to build the model was 70.7 ± 15.8 days (range 30-90) and the period used to test the model was 44.1 ± 21.0 days (range 16-90).

Results

Characteristics of the study population

In all, 29 patients were recruited, but only 21 (72%) completed the study. Reasons for drop-out were: unplanned moving (n = 2) and too short period of registration (< 20 days of registration) in 2010 (n = 5) and in 2011 (n = 1). The characteristics of the patients completing the study are reported in **table 2**. In most of

the patients, allergic rhinitis had started 3 or more years before (average disease duration 6.1 ± 0.8 years). The average age at disease onset was 4.6 ± 0.5 years. Overall, 12/21 patients had also bronchial asthma. All the patients had serum total IgE levels > 150 kU/l and specific IgE antibodies to at least one of the tested major pollen allergenic proteins (**table 2**).

Observation period and missing values

During the year used to build the model, a higher mean number of days with ACS > 0.5 (49.6 ± 22.9 vs. 22.7 ± 21.4) (t-test; p < 0.001) was observed (**table 3**).

Performance of the three predictive models

The predictive performance of the models developed on the three datasets (DMC, DMCP, DMCPI) progressively improved with the dataset size (**table 4**). The best performing dataset (DMCPI) correctly classified 77.8% ± 10.3% and 75.5% ± 13.2% (p 0.21) of the recorded days in the model and test years, respectively. In this predicting approach, 9/21 patients (42.9%) showed ≥ 80% correct classification of the recorded days in both years. The **figure 2** shows the models' performance using the DMCPI dataset. On the DMCPI dataset, the mean sensitivity and specificity parameters, for both model and test years, resulted to be high (78.2 ± 13.4 and 74.9 ± 14.2 respectively).

Figure 2 - *Mean percentage of correct classification for each patient for both model and test sets using the DMCPI dataset.*
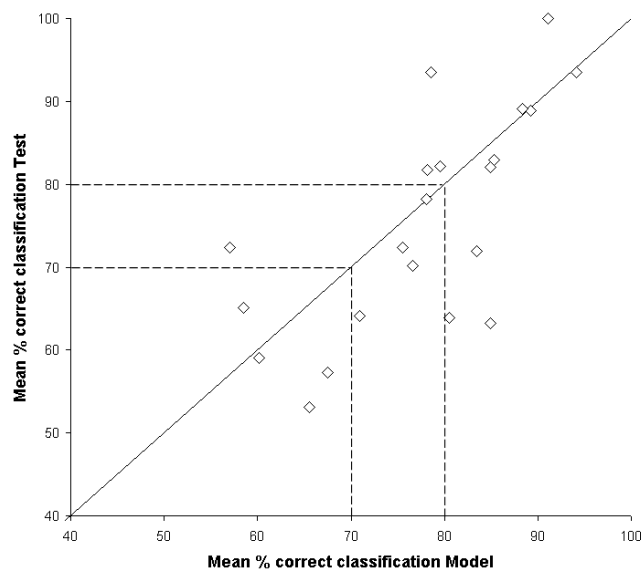
The model year performance seems to be related to that of the test year. Data are normally distributed (Shapiro-Wilk p > 0.05). The

Table 2 - Characteristics of the study population°.

	mean	SE	n	%
Age (years)	11.7	0.7		
Allergic rhinitis				
Age of onset (y)	4.6	0.5		
Duration (y)	6.1	0.8		
ARIA classification (severity)§				
mild				
intermittent (n,%)			1	4.8%
persistent (n,%)			5	23.8%
moderate/severe				
intermittent (n,%)			8	38.1%
persistent (n,%)			7	33.3%
Asthma				
Age of onset (y)	4.8	0.8		
Duration (y)	4.7	0.9		
GINA classification (severity)§*				
absent (n,%)			9	42.8%
intermittent (n,%)			10	47.6%
persistent mild (n,%)			1	4.8%
persistent moderate/severe (n,%)			1	4.8%
IgE responses*				
Total IgE (kU/l)*	499.8	89.0		
rBet v 1**	2.3	1.5		23.8%
rPhlp1**	29.8	4.6		95.2%
rPhlp5b**	34.1	20.1		76.2%
nOle e 1**	6.2	14.7		76.2%
rPar j 2**	37.5	18.6		23.8%
sum IgE-SA*	0.1	0.0		

° 21 participants, 12 (57.1%) males
§ n and % of patients with moderate/severe symptoms (criterion to classify the severity of allergic rhinitis - AR)
§* n and % of patients with persistent moderate/severe symptoms (criterion to classify the severity of allergic rhinitis - AR)
* geometric means and standard errors
** geometric means and standard errors on positive values

best performing patients tended to a higher degree of atopy, i.e. to a higher number of pollen sensitizations (figure 2,3). Similarly, poor prediction performances (< 70%) were observed among the patients with a lower degree of atopy, characterized by IgE-SA < 0.15 (data not shown). A time lag > 1 tended to be associated with better performances and no significant differences were observed between model and test year at each time lag (**table 5**). In **figure 4** the forecasting results of each single patient's PLSDA model at time lag 4 on the DMCPI dataset during the month of May of the testing year were reported. It is possible to observe the high accuracy of each patient specific model and the different patient-specific day of passing the threshold value.

C. Costa, P. Menesatti, M.A. Brighetti, A. Travaglini, V. Rimatori, A. Di Rienzo Businco, S. Pelosi, A. Bianchi, P.M. Matricardi, S. Tripodi

Figure 3 - *Mean percentages of correct classification for each patient for both model and test sets using the DMCPI dataset. Error bars indicate SE. t-test was used to compare model and test percentages (* p < 0.05).*
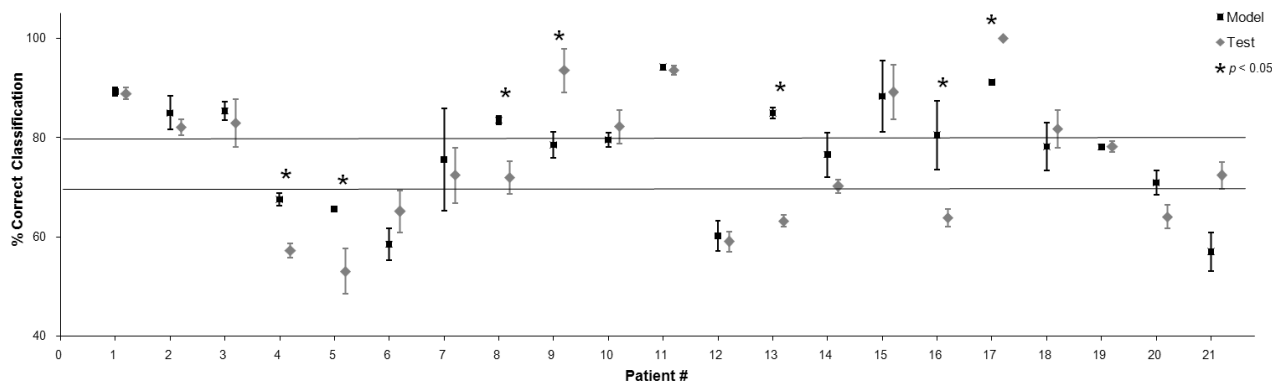

Table 3 - *Recorded data during the Model and the Test period.*

	Model		Test		p°
	mean	SD	mean	SD	
Consecutive recorded days	70.7	15.8	44.1	21.0	> 0.0001
Days with symptoms*	49.6	22.9	22.7	21.4	> 0.0001

* a cut-off of ACS> 0.5 has been used for positivity
° t-test comparing model and test percentages

In the upper left box (A), the input datasets (x-block) composed by the 15 variables were represented by 4 different shapes identifying the 4 different groups of variables: ordinal dates (1 variable), meteoclimatic (5 variables), pollen concentrations (5 variables) and IgE-pc (4 variables). In the upper right box (B) the response variable (y-block) was represented as a cross shape. The central box (C) summarizes the PLSDA model building and selection using, for each patient, the model year dataset; the x-block was preprocessed using an autoscale algorithm, then a second row preprocessing step was applied using seven different algorithms: none, baseline (Weighted Least Squares), detrend (remove a linear trend), mean centering, msc (multiplicative scatter correction with offset), normalize (normalization of the rows) and snv (Standard Normal Deviate). For each dataset, the best models were extracted at 5 different forecast levels (time lag) ranging from 0 to 4 days and 15 Latent vectors (from 1 up to 15). This leads to a total of 4,410 (210 for each patient) potential models for dataset DMC, 8,085 (385 for each patient) potential models for

Table 4 - *Performances of the best algorithm, by database used for calculation, in 21 patients with hay fever.*

	DMC			DMCP			DMCPI		
	Model	Test	p°	Model	Test	p°	Model	Test	p°
Average performance (mean, SD)	70.1	54.8	> 0.0001	76.1	66.2	> 0.0001	77.8	75.5	0,035
	16,8	17.5		8.9	8.7		10.3	13.2	
# Patients ≥ 80% pcc* (%)	38	5		38	5		43	43	
70 ≤ # Patients < 80% pcc* (%)	5	19		38	38		33	24	
# Patients ≤ 70% pcc* (%)	57	76		24	57		24	33	

°t-test comparing model and test percentages
*pcc: percentage of correct classification

Figure 4 - *Forecasting results of the single patient's PLSDA model at time lag 4 on the DMCPI dataset during the month of May of the test year.*
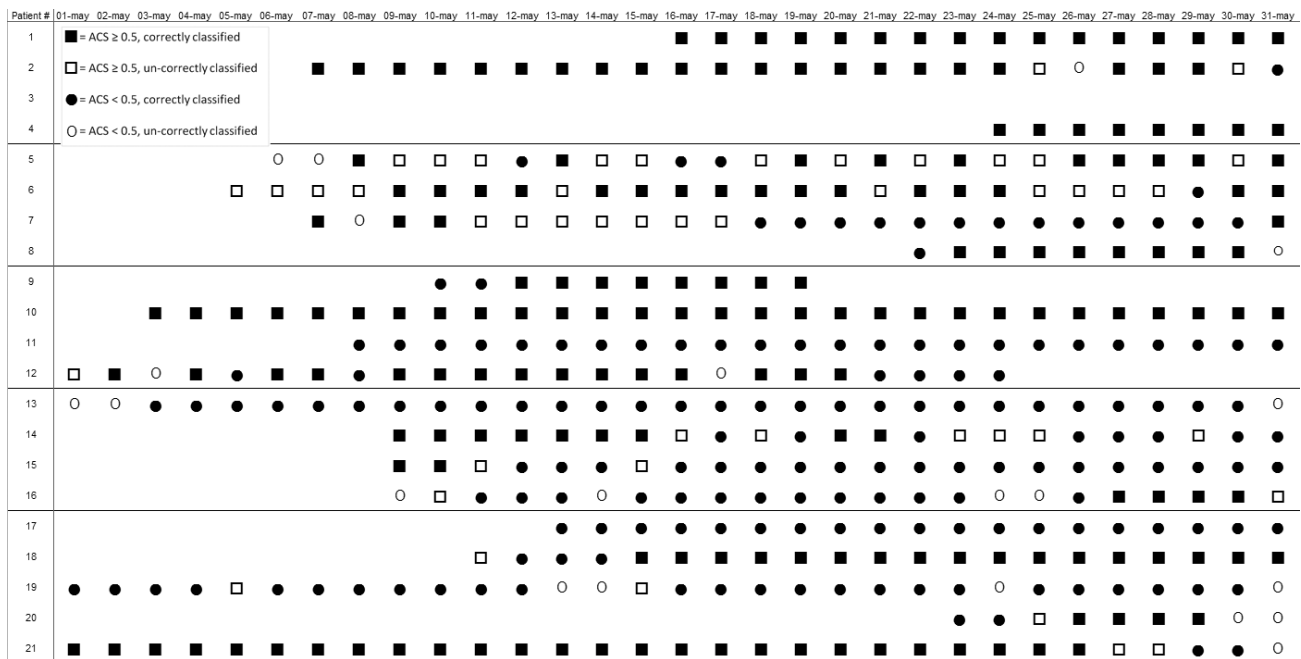
Black squares = ACS ≥ 0.5, correctly classified; white squares = ACS ≥ 0.5, un-correctly classified; black circles = ACS < 0.5, correctly classified; white circles = ACS < 0.5, un-correctly classified.

dataset DMCP and 11,025 (525 for each patient; reported in the scheme) potential models for dataset DMCPI to be elaborated. The application of the selected models on the second year dataset was summarized in the box D using the different shapes used in box A. The input dataset in box D, applying the model selected in box C, extracted for each time lag, reports a provisional output of the ACS values as below or above 0.5 (box E).

Table 5 - *Influence of the time lag on the algorithm prediction performance in 21 patients with hay fever (DMCPI dataset).*

	Model pcc*		Test pcc*		p
	mean	SD	mean	SD	
0	78.3	12.9	73.1	14.7	0.23
1	75.8	12.1	72.8	13.8	0.45
2	80.4	14.6	77.5	15.1	0.53
3	78.2	79.1	79.1	13.1	0.60
4	76.9	12.3	75.3	16.3	0.72

* pcc: percentage of correct classification
° t-test

Discussion

In a prospective study of children with hay fever, we tested the symptom predictive accuracy of forecast models of increasing complexity. We found that a multivariate modeling approach can accurately predict the presence or absence of symptoms up to 4 days before the event. We also found that the models' predictive performance tended to improve when the degree of individual allergic susceptibility was also taken into account. Finally, we found that the predictive performance improves at time lag values > 1 day after exposure.

The performance obtained by our model is relatively good, and comparable to the one obtained in previous studies. The symptoms of seasonal allergic rhinitis in 102 adult Austrian patients were recently predicted with a model based on day number of the year, grass pollen counts of the previous 2 weeks, forecasted grass pollen counts, maximum and mean temperatures (7). This model had a similar performance (76%) as the one (78%) obtained in our Italian children. Similarly, de Weger et al. (33) predicted the severity of symptoms in 80 (Netherlands) adults with hay fever with a model based on three risk categories and a time-lag of 1-5 days. The prediction performance ranged between 65% and 77% (33). Interestingly, our model - although

tested in areas with higher aerobiological complexity and in highly polysensitized patients - provided similar performances as the Austrian and Netherlands models.

The reasons of such a good performance deserve further discussion. First, our model has some strength in its statistical approach, when compared to the previous ones: it was tested by comparing data acquired in different years, rather than within the same years. The use of an external control (model vs. tested year) is considered more robust (34,15,16) and can increase the reliability of the prediction. Second, our model was tailored on the patient's sensitization profile detailed at molecular level thanks to IgE testing, while other models have been considering only patient-independent environmental components. A personalized forecast of symptoms has been recently proposed (35), but it was based on the patient' threshold of sensitivity (i.e. a subjective parameter). Conversely, the evaluation of patient's susceptibility is in our model based on the objective evaluation of his/her sensitization profile, detailed at molecular level and related to the total IgE levels (30,35,36). Our model can therefore strengthen or weaken the impact of the pollen count variables on the basis of their clinical relevance in the individual patient. Third, the best performance was obtained by using a > 2 days lag time between pollen counts and patient's symptoms. This observation is in agreement with previous studies (33), the widely accepted concept of a delay between meteorological modifications and their consequences on pollination (16) and a delay or cumulative impact of pollen exposure on symptoms (13).

Our approach is a good basis for further developments. The proposed method has been engineered using a partitioning based on the 2 years of recorded data by each patient. An adaptive routine, based on the same PLSDA multivariate approach, may auto-train and update the model at each new daily record, thus adding new data to the historical ones. The possibility to have different time lag models will also allow forecasting symptoms up to one week. Patients often do not follow a daily therapy, and an early prevision of symptoms could improve the adherence. This kind of approach, when automatically integrated with pollen observatory and meteo-climatic stations, could be implemented on mobile devices to return the patient online feedback. Therefore, the allergic symptoms forecast may aid allergy sufferers to avoid exposure to atmospheric concentrations of allergenic pollen, and help them plan taking medication, always under medical supervision (37). Hay fever is not a disease severe enough to generate excessive anxiety and, moreover, is one of those diseases whose level of symptoms can be modulated by changing behavior (e.g. by improving adherence to medication or by reducing allergen exposure). We therefore believe that predicting symptoms can improve the patient's self-confidence and disease self-management.

We have to acknowledge some limitation in our study. First, that the population sample examined in this pilot study is relatively small and that our conclusions would have been more solid if based on a larger dataset in the next future. Moreover, the informatics platform used in our study runs, partially, on a normal computer, an approach that may be considered old fashioned. This platform is however now available on smart phones. Third, patients with less than 20 recording days were not included in the analysis, but these patients have usually a lower compliance and a worse symptoms control. In a future larger sized study, our prediction tool should be evaluated also for these patients. Fourth, the population setting included only children and the conclusion of this study need to be reproduced in an adult population. Similarly, the generalizability of our conclusions is geographically limited, and studies with the same approach should be done in area with different climatic and aerobiological conditions. Moreover, the ACS threshold (0.5) values, which indicates moderate symptoms, could seem of less importance for patient life, but *i.* we observed different patient-specific days of passing the threshold value (**figure 4**), indicating a model behavior tailored on each patient and *ii.* the possibility to use other ACS threshold values for more severe symptoms, when the dataset size will be increased (in this study too few patients showed severe symptoms). The proposed approach could be applied to other Medication Scores.

In conclusions, this monocentric study in a small population shows that symptom forecasts of seasonal allergic rhinitis is possible also in highly polysensitised patients in geographic areas with complex pollen exposure, provided that predicting models are made precise as possible by tailoring their algorithms to the individual patient's allergic susceptibility. Future studies will have to monitor how e-diaries and predictive algorithms can influence adherence to treatment, at the extremes and during the peak of the pollen season. Multicenter studies in large population samples adopting the same acquisition data system on smart phones are now needed to confirm and reinforce this encouraging conclusion.

References

1. Asher MI, Montefort S, Björkstén B, et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. Lancet. 2006;368:733-43.

2. Brozek JL, Bousquet J, Baena-Cagnani CE, et al. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 Revision. Journal of Allergy and Clinical Immunology. 2010;126:466-76.

3. Canonica GW, Bousquet J, Mullol J, Scadding GK, Virchow JC. A survey of the burden of allergic rhinitis in Europe. Allergy. 2007;62(suppl 85):17-25.

4. Emberlin J, Mullins J, Cordon J, et al. Regional variations in grass pollen seasons in the UK, long term trends and forecast models. Clinical and Experimental Allergy. 1999;29:347-56.

5. Peden D, Reed CE. Environmental and occupational allergies. Journal of Allergy and Clinical Immunology. 2010;125:S150-60.

6. Traidl-Hoffmann C, Jakob T, Behrendt H. Determinants of allergenicity. Journal of Allergy and Clinical Immunology. 2009;123:558-66.

7. Voukantsis D, Karatzas K, Jaeger S, Berger U, Smith M. Analysis and forecasting of airborne pollen–induced symptoms with the aid of computational intelligence methods. Aerobiologia. 2013;29:175-85.

8. Bousquet J, Khaltaev N, Cruz AA, et al. Allergic Rhinitis and its Impact on Asthma (ARIA) 2008 update (in collaboration with the World Health Organization, GA(2)LEN and AllerGen). Allergy. 2008;63(Suppl 86):8-160.

9. Häfner D, Reich K, Matricardi PM, Meyer H, Kettner J, Narkus A. Prospective validation of 'Allergy-Control-SCORETM': a novel symptom–medication score for clinical trials. Allergy. 2011;66:629-36.

10. Berger U, Karatzas K, Jaeger S, Voukantsis D, Sofiev M, Brandt O, et al. Personalized pollen−related symptom−forecast information services for allergic rhinitis patients in Europe. Allergy. 2013;68:963-5.

11. D'Amato G, Cecchi L, Bonini S, et al. Allergenic pollen and pollen allergy in Europe. Allergy. 2007;62:976-90.

12. Moreno-Grau S, Elvira-Rendueles B, Moreno J, et al. Correlation between Olea europaea and Parietaria judaica pollen counts and quantification of their major allergens Ole e 1 and Par j 1-Par j 2. Annals of Allergy, Asthma & Immunology. 2006;96:858-64.

13. Feo Brito F, Mur Gimeno P, Carnes J, et al. Olea europaea pollen counts and aeroallergen levels predict clinical symptoms in patients allergic to olive pollen. Annals of Allergy, Asthma & Immunology. 2011;106:146-52.

14. Astel A. Chemometrics based on fuzzy logic principles in environmental studies. Talanta. 2007;72:1-12.

15. Menesatti P, Antonucci F, Costa C, Mandalà C, Battaglia V, La Torre A. Multivariate forecasting model to optimize management of grape downy mildew control. Vitis. 2013;52:141-8.

16. Brighetti MA, Costa C, Menesatti P, Antonucci F, Tripodi S, Travaglini A. Multivariate statistical forecasting modelling to predict Poaceae pollen critical concentrations by meteoclimatic data. Aerobiologia. 2014;30:25-33.

17. Abraham B, Ledolter J. (2009). Statistical methods for forecasting. John Wiley & Sons, 2009.

18. Doganis P, Alexandridis A, Patrinos P, Sarimveis H. Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. Journal of Food Engineering. 2006;75:196-204.

19. Travaglini A, Mazzitelli A, Leonardi D, Grilli Caiola M. 2002. Monitoraggio di polline urbano e allergie. In: Atti dell'Accademia dei Lincei Conferenza annuale della Ricerca: Ecosistemi urbani 2002;1-40. ISBN/ISSN:0391-805X.

20. Dondi A, Tripodi S, Panetta V, et al. Pollen-induced allergic rhinitis in 1360 Italian children: comorbidities and determinants of severity. Pediatric Allergy and Immunology. 2013;24:742-51.

21. Asher MI, Keil U, Anderson HR, et al. International study of asthma and allergies in childhood (ISAAC): rationale and methods. European Respiratory Journal. 1995;8:483-91.

22. Jáuregui I, Dávila I, Sastre J, et al. Validation of ARIA (Allergic Rhinitis and its Impact on Asthma) classification in a pediatric population: the PEDRIAL study. Pediatric Allergy and Immunology. 2011;22:388-92.

23. Bousquet J, Clark T, Hurd S, et al. GINA guidelines on asthma and beyond. Allergy. 2007;62:102-12.

24. Grouin J-M, Vicaut E, Jean-Alphonse S, et al. The average Adjusted Symptom Score, a new primary efficacy end-point for specific allergen immunotherapy trials. Clinical & Experimental Allergy. 2011;41:1282-8.

25. UNI. Qualità dell'aria - Metodo di campionamento e conteggio dei granuli pollinici e delle spore fungine aerodisperse. UNI 11008, 2004.

26. Travaglini A., Albertini R., Zieger E., Brighetti M.A., Ugolotti M., Torrigiani T. (Eds.). Manuale di gestione e qualità della rete italiana di monitoraggio in Aerobiologia R.I.M.A.®. Bologna Tipografia LEGO, 2009. ISBN 978-88-900277-1-0.

27. Hirst JM. Changes in atmospheric spore content: Diurnal periodicity and the effects of weather. Transactions of the British Micological Society. 1952;36:375-93.

28. Mandrioli P. Metodica di campionamento e conteggio dei granuli pollinici e delle spore fungine aerodisperse. In Regione Emilia Romagna (Ed.) Il monitoraggio aerobiologico in Emilia Romagna. 1994;30:9-19.

29. Stringari G, Tripodi S, Caffarelli, et al. The impact of component resolved diagnosis on specific immunotherapy prescription in children. Journal of Allergy and Clinical Immunology. 2014;134:75-81

30. Di Lorenzo G, Mansueto P, Pacor ML, et al. Evaluation of serum s-IgE/total IgE ratio in predicting clinical response to allergen-specific immunotherapy. Journal of Allergy and Clinical Immunology. 2009;123:1103-10.

31. Sjöström M, Wold S, Söderström B. PLS Discriminant Plots, Proceedings of PARC in Practice. Elsevier Science Publishers B.V. North-Holland, 1986.

32. Costa C, Antonucci F, Boglione C, Menesatti P, Vandeputte M, Chatain B. Automated sorting for size, sex and skeletal anomalies of cultured seabass using external shape analysis. Aquacultural Engineering. 2013;52:58-64.

33. de Weger LA, Beerthuizen T, Hiemstra PS, Sont JK. Development and validation of a 5-day-ahead hay fever forecast for patients with grass-pollen-induced allergic rhinitis. International Journal of Biometeorology. 2013;1-9.

34. Adams-Groom B, Emberlin J, Corden J, Millington W, Mullins J. Predicting the start of the birch pollen season at London, Derby and Cardiff, United Kingdom, using a multiple regression model, based on data from 1987 to 1997. Aerobiologia. 2002;18:117-23.

35. Christensen LH, Holm J, Lund G, Riise E, Lund K. Several distinct properties of the IgE repertoire determine effector cell degranulation in response to allergen challenge. Journal of Allergy and Clinical Immunology. 2008;122:298-304.

36. Hamilton RG, MacGlashan DWJr, Saini SS. IgE antibody-specific activity in human allergic disease. Immunologic Research. 2010;47:273-84.

37. Berger U, Karatzas K, Jaeger S, et al. Personalized pollen-related symptom-forecast information services for allergic rhinitis patients in Europe. Allergy. 2013;68:963-5.