

Designing Predictive Models for Beta-Lactam Allergy Using the Drug Allergy and Hypersensitivity Database



Anca Mirela Chiriac, MD^{a,b}, Youna Wang, MD, PhD^{a,c}, Rik Schrijvers, MD, PhD^{a,d}, Philippe Jean Bousquet, MD, PhD^a, Thibault Mura, MD, PhD^{e,f}, Nicolas Molinari, PhD^{e,f,g}, and Pascal Demoly, MD, PhD^{a,b} Montpellier and Paris, France; Wuhan, China; and Leuven, Belgium

What is already known about this topic? Drug allergy evaluation for suspected beta-lactam allergy relies on allergy testing (usually *in vivo* tests, i.e., skin tests and drug provocation tests).

What does this article add to our knowledge? This study assessed in a large cohort the value of 2 independent prognostic methods (based on clinical history predictors) in the diagnosis of beta-lactam allergy.

How does this study impact current management guidelines? Both algorithms demonstrated poor sensitivity, demonstrating that history alone could not replace a conventional allergy evaluation for suspected beta-lactam allergy.

BACKGROUND: Beta-lactam antibiotics represent the main cause of allergic reactions to drugs, inducing both immediate and nonimmediate allergies. The diagnosis is well established, usually based on skin tests and drug provocation tests, but cumbersome. **OBJECTIVES:** To design predictive models for the diagnosis of beta-lactam allergy, based on the clinical history of patients with suspicions of allergic reactions to beta-lactams.

METHODS: The study included a retrospective phase, in which records of patients explored for a suspicion of beta-lactam allergy (in the Allergy Unit of the University Hospital of Montpellier between September 1996 and September 2012) were used to construct predictive models based on a logistic regression and decision tree method; a prospective phase, in which we performed an external validation of the chosen models in patients with suspicion of beta-lactam allergy recruited from 3 allergy centers (Montpellier, Nîmes, Narbonne) between March and November 2013. Data related to clinical history and allergy evaluation results were retrieved and analyzed.

RESULTS: The retrospective and prospective phases included 1991 and 200 patients, respectively, with a different prevalence of confirmed beta-lactam allergy (23.6% vs 31%, $P = .02$). For the logistic regression method, performances of the models were similar in both samples: sensitivity was 51% (vs 60%), specificity 75% (vs 80%), positive predictive value 40% (vs 57%), and negative predictive value 83% (vs 82%). The decision tree method reached a sensitivity of 29.5% (vs 43.5%), specificity of 96.4% (vs 94.9%), positive predictive value of 71.6% (vs 79.4%), and negative predictive value of 81.6% (vs 81.3%).

CONCLUSIONS: Two different independent methods using clinical history predictors were unable to accurately predict beta-lactam allergy and replace a conventional allergy evaluation for suspected beta-lactam allergy. © 2017 American Academy of Allergy, Asthma & Immunology (J Allergy Clin Immunol Pract 2018;6:139-48)

Key words: Allergy testing; Beta-lactams; Drug hypersensitivity; Predictive models

^aDepartment of Pulmonology, Division of Allergy, Hôpital Arnaud de Villeneuve, University Hospital of Montpellier, Montpellier, France

^bUPMC Univ Paris 06, Sorbonne Universités, Paris, France

^cDepartment of Allergy, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

^dLaboratory of Clinical Immunology, Department of Microbiology and Immunology, KU Leuven, Leuven, Belgium

^eDepartment of Clinical Information, Clinical Research Unit, University Hospital of Montpellier, Montpellier, France

^fMontpellier University, Montpellier, France

^gIMAG, UMR 5149, University Hospital of Montpellier, Montpellier, France

This study was funded by the University Hospital of Montpellier, via the *Contrat Fort Chercheur* Funds and by the *COBIOPHAD project* (Horizon2020 project, supported by the European Commission, grant no. 688448).

Conflicts of interest: A.-M. Chiriac has received research support from COBIOPHAD project (Horizon2020 project, supported by the European Commission, grant no. 688448). P. Demoly is on the boards for ALK, Stallergenes Greer, ThermoFisher Scientific, Chiesi, and Menarini; has received research support from AstraZeneca; and has received lecture fees from ALK, Stallergenes Greer, and Meda. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication November 1, 2016; revised March 26, 2017; accepted for publication April 17, 2017.

Available online July 21, 2017.

Corresponding author: Anca Mirela Chiriac, MD, Allergy Unit, Arnaud de Villeneuve Hospital, University Hospital of Montpellier, 371, Avenue du Doyen Gaston Giraud, 34295 Montpellier Cedex 5, Montpellier, France. E-mail: a-chiriac@chu-montpellier.fr.

2213-2198

© 2017 American Academy of Allergy, Asthma & Immunology

<http://dx.doi.org/10.1016/j.jaip.2017.04.045>

*Abbreviations used**BL- beta-lactams**DAHD- Drug Allergy and Hypersensitivity Database**DH- drug hypersensitivity**DPT- drug provocation test**NPV- negative predictive value**OR- odds ratio**PPV- positive predictive value**ST- skin test**YI- Youden index*

Framingham cardiovascular risk score¹ and the Acute Physiology And Chronic Health Evaluation [APACHE] score²) have made their way into clinical practice, because their impact in clinical settings was shown to improve practitioners' decision making and ultimately patient outcome.

Beta-lactam (BL) antibiotics represent the main cause of allergic reactions to drugs, inducing both immediate and non-immediate reactions (occurring <1 hour and >1 hour after the last drug intake, respectively).^{3,4} In Europe, diagnosis is based on skin tests (STs) and drug provocation tests (DPTs) using procedures endorsed by the European Network on Drug Allergy/ Drug Allergy Interest Group of the European Academy of Allergy and Clinical Immunology.³ The standardization process allows drug hypersensitivity (DH) cases to be collected in databases including patients with confirmed DH and those with a negative drug allergy evaluation. In 2007, the Drug Allergy and Hypersensitivity Database (DAHD) was created.⁵ It includes all the patients with a clinical history of DH who undergo an allergy evaluation in our unit.

In this work, we retrospectively designed predictive models for BL allergy diagnosis using patients with suspicion of BL allergy included in DAHD and assessed their performance prospectively in new patients.

METHODS

Study design and patient recruitment

The study included 2 phases: a retrospective one, with patients who served as a training sample, to construct the predictive models and a prospective one in which patients were included to perform an external validation of the chosen models.

For the retrospective phase, the DAHD was questioned to retrieve all patients who underwent drug allergy evaluation for a suspicion of BL allergy between January 1996 and September 2012 in the Allergy Unit of the University Hospital of Montpellier, France. Patients with a confirmed diagnosis (allergic or nonallergic to BL) were included in the analysis. The prospective phase was carried out in 3 tertiary allergy referral centers in the South of France (Montpellier, Nîmes, and Narbonne). Adult patients (≥ 18 years old, at the time of the drug allergy evaluation) with a suspicion of BL allergy were included prospectively, from March 2013 to November 2013. The local ethics committee "Comité de Protection des Personnes Sud-Méditerranée III" approved the design of the study. For the prospective study, all patients gave their informed written consent to take part in this research. Patients included in the retrospective study had given their consent to take part in clinical research at the time of the allergy evaluation.

Drug allergy evaluation and patient groups

Drug allergy evaluation for BL allergy diagnosis followed European Network on Drug Allergy recommendations^{3,6-8} and it was identical for all centers. A comprehensive drug allergy evaluation included collection of clinical history data, as well as *in vivo* allergy testing (STs and, in absence of contraindication, DPTs).

The predicted outcome was the diagnosis of BL allergy. A patient was considered allergic if the ST or DPT was positive for the culprit BL. A patient was considered nonallergic if both the ST and DPT were negative for the suspected BL. If a patient had suffered from reactions to different BLs, a potential allergy to a common structure, present in all the BLs involved, was inferred. If only 1 BL among different suspected BLs was tested and found to be negative, the patient was considered nonallergic.

Statistical methods

Sample size. For the retrospective phase, we included all the patients with a definite diagnosis (allergic or nonallergic). For the prospective part, the number of needed subjects was a priori calculated to obtain composite criteria of allergy with a sensitivity of 90% $\pm$ 10% based on a 35% incidence of allergy (considered from our previous reports and the reports of others). With these criteria, 100 patients were needed. We decided to include 200 patients in the prospective cohort to validate the model in the original cohort.

Collection of patients' data. DAHD is based on FileMakerPro and is approved by the French data protection agency. It collects data on 8 clinical history variables: general patient data (sex, atopy, asthma) and data specifically related to the reaction: age (at the time of reaction), semiology (signs and symptoms of the reaction), delay (between the reaction and the allergy evaluation); chronology (time to reaction onset after the last drug intake), and the BL involved. The 9 variables (the aforementioned 8 plus the number of clinical histories related to BL/patient) included in the analysis are available in [Appendix E1](#) in this article's Online Repository at www.jaci-inpractice.org. For patients with multiple reactions, the most severe reaction was taken into account (chronologically, this was typically the most recent one).

Descriptive analysis. Data were expressed either as numbers and percentage (with 95% CIs) or as medians (with interquartile range). Comparisons were made using chi-square tests (or Fisher exact tests for small sample sizes) for categorical variables (ie, sex, atopy, asthma, BL involved, semiology, chronology, age at reaction [in classes], number of clinical histories related to BL/patient, and patient's profile according to the number and type of BL involved in the clinical histories [[Table 1](#)]) and Wilcoxon Mann-Whitney tests (in case of nonnormal distribution of continuous variables, ie, age at reaction [in years] and delay).

Risk factor analysis. Clinical risk factors were searched using univariate logistic regression. To appreciate independent clinical risk factors linked to allergy diagnosis, a multivariate analysis was performed, enabling the calculation of odds ratios (ORs) and 95% CIs. Variables were selected if the *P* value was less than .15 in the univariate analysis, and a forward procedure was used to select the final model. Interactions between variables were tested. We chose the model that minimized the Akaike information criterion.⁹

Construction of the predictive score by logistic regression. The predictive score was built on the retrospective sample. Our objective was to construct a predictive model. To

TABLE I. Descriptive analysis of the patients from the training set

Clinical characteristics of retrospective patients	Total patients N = 1991	Confirmed allergy N (%) N = 470	P
Sex*			.69†
Female	1360	324 (23.8)	
Male	621	143 (23.0)	
Asthma			.75†
Asthmatic	1741	409 (23.5)	
Nonasthmatic	250	61 (24.4)	
Atopy			.93†
Atopic	1081	256 (23.7)	
Nonatopic	910	214 (23.5)	
Antibiotic			.14†
Penicillins	1535	351 (22.9)	
Cephalosporins	432	110 (25.4)	
Other BL	24	9 (37.5)	
Chronology of the reaction*			<.001‡
<1 h	397	183 (46.1)	
1-6 h	236	55 (23.3)	
6-24 h	218	49 (22.5)	
>24 h	686	128 (18.6)	
Unknown	439	55 (11.7)	
Semiology of the reaction			<.001‡
Anaphylaxis (w/o shock)	265	100 (37.7)	
Anaphylactic shock	149	95 (63.7)	
Urticaria/angioedema	866	151 (17.4)	
Maculopapular exanthema	472	77 (16.4)	
Isolated respiratory symptoms	24	4 (16.6)	
Other cutaneous symptoms	29	4 (13.8)	
Other symptoms	177	37 (21)	
Unknown	9	2 (22.2)	
Age at reaction			<.001‡
Child (<18 y)	718	90 (12.5)	
Adult (≥18 y)	1273	380 (29.8)	
Number of reactions/patient			<.001‡
1 episode	1619	340 (21.0)	
Multiple episodes	372	130 (35.0)	
Patient's profile according to the number and type of the BL involved			<.001‡
1 episode (1 BL involved in only 1 episode)	1619	340 (21.0)	
Single reactor (1 BL involved in ≥2 episodes)	138	50 (36.2)	
Multiple reactor (≥2 involved in ≥2 episodes)	234	80 (34.2)	
Delay (mo), median (Q 25-75)	52.26 (7.06-183)	84.1	<.001‡

Bold values indicate statistical significance ($P < .05$).

*Variables containing missing values (≤ 15 missing values/variable, ie, $\leq 1\%$ per variable).

†Fisher exact or χ^2 tests for %.

‡Wilcoxon test for other data.

establish the composite score, we gave a score to each variable included in the final prediction model in relation to each one's β parameter (regression coefficient) in that model.¹⁰ To obtain the total score, we added the scores of each variable selected and a receiver operating characteristic curve was established to estimate the area under the curve of the score. The cutoff score that maximized the Youden index ($YI = \text{Sensitivity} + \text{Specificity} - 1$) was chosen. The discriminative ability of the score was estimated in terms of sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).¹¹

Validation of the logistic regression model. An external validation was carried out on the prospective sample.¹² The performance of the model was checked by calibration (by means of the Hosmer-Lemeshow test, by plotting the observed proportions of events, ie, allergy to BL, against the predicted probabilities for groups defined by ranges of predicted risk) and discrimination (by using receiver operating characteristic curves with calculation of the area under the curve). The diagnostic value of the model was calculated in terms of sensitivity, specificity, and NPVs and PPVs, with approximate 95% CI.

Construction of the decision tree. The retrospective sample was used to define the decision tree, using the Classification and Regression Tree *method* (package *rpart*).¹³ Decision trees are represented by a set of questions that split the learning sample into smaller and smaller parts, called nodes. Classification and Regression Tree asks only yes/no questions. Classification and Regression Tree algorithm searches for all possible variables and all possible values to find the best split (ie, to ideally have “pure” terminal nodes, where 1 class dominates, eg, allergic/nonallergic). All the 9 variables were introduced in the model (input), but only 4 variables were selected by *rpart* to build the decision tree (output) and predict allergy occurrence (semiology, delay, chronology, age), the most discriminative one being semiology. The optimal tree size was determined by cross-validation. The discriminative abilities of the model (sensitivity, specificity, PPV, NPV) were estimated. The constructed tree can then be used to classify new, unseen observations.

Validation of the decision tree. The validation of the tree algorithm was done on the prospective sample, by the estimation of the rate of misclassification of allergic and nonallergic patients.

Missing data and miscellaneous. Because of the low number of missing data ($\leq 1\%$), no imputation method was used. All directional *P* values were 2-tailed, and significance was assigned to values less than .05. The statistical analysis was performed with the help of SAS software (version 9.3, SAS Institute Inc, Cary, NC) and R (R-3.2.3. pkg).

RESULTS

Study population

During the study period, 2543 patients (2343 in the retrospective sample and 200 in the prospective one) underwent a drug allergy evaluation for suspicion of BL allergy. The following patients were excluded from analysis of the retrospective sample: 143 patients who had incomplete data in DAHD (the name of the suspected drug class motivating the BL allergy evaluation was not available), 209 patients who had an incomplete drug allergy evaluation (200 patients lost to follow-up before completing the allergy testing and 9 patients with negative ST result and a contraindication for DPT, ie, patients presenting with severe cutaneous adverse reactions). Thus, 1991 patients were analyzed within the retrospective sample.

Regarding the prospective sample, only 7 patients refused the participation in the study. From the 200 patients included in the analysis, 135 were recruited in Montpellier, 50 in Nîmes, and 15 in Narbonne.

Descriptive analysis of the training set

The clinical characteristics of the patients in the retrospective sample are summarized in Table I. Of the 1991 patients, 470 (23.6%) were allergic to the suspected BL. Female patients accounted for two-third of the total sample. The patients reported mostly nonimmediate reactions (80%) and penicillins were the most frequently involved BL (77%). The most frequent clinical entity was urticaria/angioedema (866 patients [43.5%]), followed by maculopapular exanthema (471 patients [23.7%]) and anaphylaxis (with or without [w/o] shock, 414 [20.7%]). Less than 10% of the patients reported vaguely recalled symptoms and/or could not give any information about the event that had labeled them as “allergic.” At the time of the allegedly allergic event, two-third of the patients were adults. One in 5 patients

(372 [18.8%]) reported multiple episodes, including either the same BL (single reactors) or several different BLs (multiple reactors) (138 patients [36.2%] and 234 patients [34.2%], respectively). The median delay between the (last) allergic event and the drug allergy evaluation was 52.2 months (range, 1 month to 50 years), and a quarter of the patients were tested more than 15 years after the last reaction. One-third of the patients did not recall the name of the culprit BL. There were significantly more allergic patients in the immediate reactors’ group versus nonimmediate reactors ($P < .001$), in the anaphylaxis group (with or w/o shock) compared with all other clinical manifestations ($P < .001$), and among adults versus children ($P < .001$). In addition, patients reporting 2 or more allergic events were more often allergic than those reporting 1 allergic event ($P < .001$) and the mean delay from reaction to allergy evaluation was shorter in the allergic group compared with the nonallergic patients (84.1 months vs 128.8 months, respectively; $P < .001$). There were no differences between allergic and nonallergic subjects in terms of sex ($P = .7$), antibiotic involved in the allergic events ($P = .14$), or comorbidities such as asthma ($P = .75$) and atopy ($P = .93$).

Search for risk factors in the training set

ORs and adjusted OR for univariate and multivariate analysis conducted on the training set are presented in Table II. In univariate analysis, sex, asthma, atopy, and the antibiotic involved were not associated with the allergic outcome. Possible risk factors for allergy were identified: a shorter delay between the reaction and the allergy evaluation, immediate reactions, adult age at the time of the reaction, anaphylactic reactions, and multiple episodes. After adjustment on each of the variables considered in the model, the delay reaction/allergy evaluation became nonsignificant (OR = 0.99; 95% CI, 0.99-1; $P = .14$) and penicillin became significant. All the other risk factors identified in univariate analysis remained significantly associated with allergy. Thus, patients with immediate reactions or anaphylaxis had a 2 and 2 to 4 time higher risk, respectively, to be allergic, compared with those with nonimmediate reactions or reporting other clinical entities. Similarly, if the event involved a penicillin rather than another BL, the risk of being allergic was 1.53 times higher (95% CI, 1.15-2.04; $P = .005$). Conversely, the occurrence of the reaction during childhood (< 18 years) reduced the risk of it being a true allergic event (OR = 0.49; 95% CI, 0.37-0.65; $P < .001$).

Descriptive analysis of the validation set and comparison to the training set

When compared with the retrospective sample, the prospective sample showed more demonstrated drug allergies (31.0% vs 23.6%; $P = .02$), less atopic (27.0% vs 45.7%; $P < .001$), and more adult patients at the time of the reaction (80.8% vs 64.0%; $P < .0001$). There were also significantly more patients with multiple episodes in the prospective sample (24.5% vs 18.7%; $P = .04$), more immediate reactions (37.0% vs 20.1%; $P < .001$), more anaphylactic shocks (14.0% vs 7.5%; $P < .001$), and more reactions involving penicillins (88.5% vs 77.1%; $P < .001$). The characteristics of the patients in the validation sample are summarized in Table III.

The prognostic model using logistic regression

When assessing calibration on the original (training) sample, the Hosmer-Lemeshow test showed a significant difference ($P = .004$) between observed and expected number of allergic individuals.

TABLE II. Risk factors associated with BL allergy (training set)

Risk factors	Univariate analysis			Multivariate analysis		
	OR	95% CI	P	Adjusted OR	95% CI	P
Delay	0.997	0.997-0.998	<.001	0.99	0.99-1	.14
Age at reaction	1.02	1.01-1.02	<.001	1.01	1.01-1.02	<.001
Sex						
Female	1					
Male	0.95	0.76-1.19	.69			
Asthma						
Asthmatic	1.05	0.77-1.43	.75			
Nonasthmatic	1					
Atopy						
Nonatopics	1					
Atopics	0.99	0.80-1.22	.93			
Antibiotic*						
All other BLs	1					
Penicillins	0.84	0.66-1.06	.15	1.53	1.15-2.04	.005
Chronology*						
≥1 h	1					
<1 h	3.85	3.04-4.87	<.001	2.08	1.57-2.76	<.001
Semiology*						
Other reactions	1					
Anaphylactic shock	8.32	8.81-11.91	<.001	4.15	2.72-6.33	<.001
Anaphylaxis w/o shock	2.87	2.16-3.79	<.001	2.23	1.63-3.04	<.001
Age at reaction (classes)						
Adult (≥18 y)	1					
Child (<18 y)	0.33	0.26-0.43	<.001	0.49	0.37-0.65	<.001
Number of reactions/patient						
1 episode	1					
Multiple episodes	2.02	1.58-2.58	<.001	2.22	1.69-2.93	<.001

Bold values indicate statistical significance ($P < .05$).

*After regrouping categories of the same variable.

The area under the curve was 0.67 (95% CI, 0.64-0.69). Thus, the model showed poor calibration and discrimination. It showed better accuracy and predictive performance when applied to the prospective sample (Hosmer-Lemeshow test, $P = .43$; area under the curve = 0.73 [95% CI, 0.65-0.81]) (Figure 1).

The risk score to predict occurrence of allergy to BL was calculated as follows:

$$Z = -2.6541 + 1.8660 \times (\text{anaphylactic shock}) + 0.8500 \times (\text{anaphylaxis w/o shock}) + 0.0166 \times (\text{age}) + 0.8002 \times (\text{number of reactions/patient}) + 1.0239 \times (\text{chronology}) - 0.0041 \times (\text{delay} \times \text{chronology}) + 0.4285 \times \text{antibiotic} - 0.0004 \times \text{delay} \times (\text{anaphylaxis w/o shock}) - 0.0127 \times \text{delay} \times (\text{anaphylactic shock}).$$

The coding of variables in the risk score was done as follows:

- anaphylactic shock and anaphylaxis w/o shock are coded 0 for “no” and 1 for “yes”;
- number of reactions/patient is coded 0 for “1 episode” and 1 for “multiple episodes”;
- age, delay are continuous;
- chronology is 0 for “nonimmediate” and 1 for “immediate”;
- antibiotic is 0 for “other BL” and 1 for “penicillins.”

The risk score that maximized the YI on the training set ($= 0.263$), ensuring best sensitivity and specificity at the same

time, was $Z = -0.58$, and on the validation sample, the optimal threshold was $Z = -0.48$. Table IV presents the performances of the logistic model for different cutoffs.

The prognostic model using the decision tree algorithm

The decision tree is presented in Figure 2 (printout in Figure E1). It had 6 leaves, and none of them was pure (ie, no complete discrimination between allergic and nonallergic individuals). Four leaves could correctly assign the diagnosis of allergic (leaf 4) or nonallergic reaction (leaves 1, 2, and 3) in three-fourth or more of the individuals. For example, the decision tree correctly assigned the diagnosis in 82.6% of nonallergic patients if clinical history was anything other than anaphylaxis or anaphylactic shock (leaf 1) and in 78.8% of nonallergic patients if they had had a history of anaphylaxis with or w/o shock but the delay between reaction and drug allergy evaluation was more than 5 years (leaf 2).

The tree model had an accuracy of 0.806 on the training set (the retrospective sample). Thus, 19.4% of the subjects were misclassified in this model. The most important misclassification occurred in the subgroup of allergic patients, where more than two-third (331 of 470 subjects [70.5%]) were false negative. Within the misclassified allergic group in the leaves predicting

TABLE III. Descriptive analysis of the patients from the validation sample

Clinical characteristics of prospective patients	Total patients N = 200	Confirmed allergy N (%)	P*
Sex			.82*
Female	143	45 (31.4)	
Male	57	17 (29.8)	
Asthma			.08*
Asthmatic	175	58 (33.1)	
Nonasthmatic	25	4 (16)	
Atopy			.04*
Atopic	146	51 (34.9)	
Nonatopic	54	11 (20.3)	
Antibiotic†			.16*
Penicillins	177	52 (29.3)	
All other BLs (only cephalosporins)	23	10 (25.4)	
Chronology of the reaction			<.001*
<1 h	74	39 (52.7)	
1-6 h	15	4 (26.6)	
6-24 h	21	5 (23.8)	
>24 h	11	3 (27.7)	
Unknown	79	11 (13.9)	
Chronology of the reaction‡			<.001*
<1 h	74	39 (52.7)	
≥1 h	126	23 (18.2)	
Semiology of the reaction			<.001*
Anaphylaxis (w/o shock)	26	16 (61.5)	
Anaphylactic shock	28	19 (67.6)	
Urticaria/angioedema	68	11 (16.1)	
Maculopapular exanthema	49	14 (28.5)	
Other symptoms	29	2 (6.9)	
Semiology of the reaction‡			<.001*
Anaphylaxis (w/o shock)	26	16 (61.5)	
Anaphylactic shock	28	19 (67.6)	
Other reactions	146	27 (18.5)	
Age at reaction‡			<.001*
Child (<18 y)	38	3 (7.8)	
Adult (≥18 y)	160	58 (36.2)	
Number of reactions/patient			.94*
1 episode	151	47 (31.1)	
Multiple episodes	49	15 (30)	
Patient's profile according to the number and type of the BL involved			<.001*
1 episode (1 BL involved in only 1 episode)	151	47 (31.1)	
Single reactor (1 BL involved in ≥2 episodes)	29	10 (34.4)	
Multiple reactor (≥2 involved in ≥2 episodes)	20	5 (25)	
Delay (mo), median (Q25-Q75)	54 (7-288)	66.7	<.001§

Bold values indicate statistical significance ($P < .05$).

*Fisher exact or χ^2 tests for %.

†After regrouping categories of the same variable.

‡Variable containing missing values (2 missing values [1%]).

§Wilcoxon test for other data.

lack of allergy, 13 patients (3.1%) had reacted with anaphylaxis with or w/o shock during allergy evaluation, during either ST or DPT.

The tree assigned the correct diagnosis to almost all the nonallergic individuals (3.6% misclassified, 55 of 1521). Overall, the decision tree's performance was as follows: sensitivity = 29.5%, specificity = 96.4%, PPV = 71.6%, NPV = 81.6%. Table E1 in this article's Online Repository at

www.jaci-inpractice.org presents the confusion matrix for the training set.

When used to make predictions on a new data set, the prospective sample, the performance of the decision tree was improved (sensitivity = 43.5%, specificity = 94.9%, PPV = 79.4%, NPV = 81.3%). However, the tree still misclassified most of the allergic individuals. Tables E2 and E3 in this article's Online Repository at www.jaci-inpractice.org

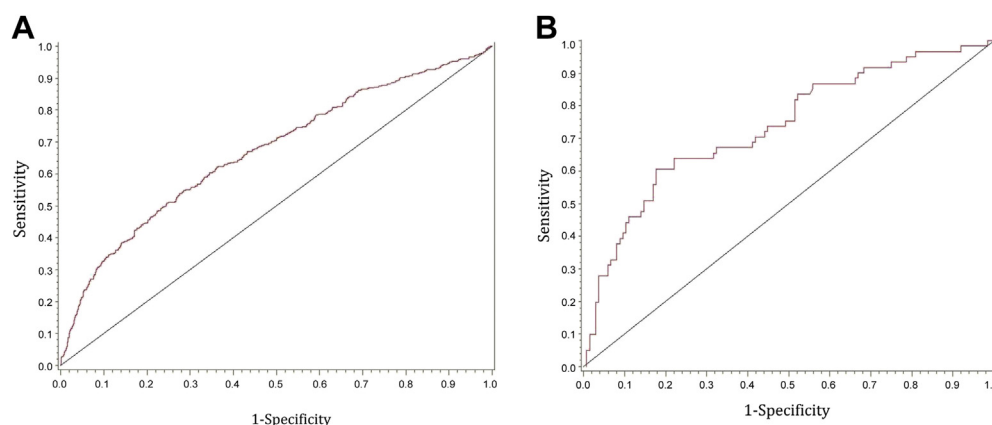


FIGURE 1. ROC curves for the retrospective (**A**) and prospective (**B**) samples. Figure 1, **A**, AUC = 0.67 (95% CI, 0.64-0.69), and **B**, AUC = 0.73 (95% CI, 0.65-0.81). AUC, Area under the curve; ROC, receiver operating characteristic.

TABLE IV. Performances of the logistic model according to different cutoffs (training, τ and validation, ν samples)

Z	−3.06	−2.10	−1.62	−1.35	−1.02	−0.58*	−0.34	−0.02	0.23	0.56
Sensitivity τ (%)	98	92	85	75	65	51*	40	30	20	10
Sensitivity ν (%)	98	93	86	75	67	60*	55	44	31	16
Specificity τ (%)	2	15	31	43	58	75*	83	92	95	98
Specificity ν (%)	4	22	41	50	60	80*	83	89	94	97
PPV τ (%)	23	25	28	29	32	40*	43	53	58	71
PPV ν (%)	31	35	40	40	42	57*	60	64	70	61
NPV τ (%)	77	86	87	84	84	83*	81	80	80	77
NPV ν (%)	85	88	87	82	80	82*	80	78	75	72

τ and ν , training and validation samples, respectively; Z, value of score.

*Corresponds to the performances of the models for the chosen cutoff (ie, the YI) on the retrospective sample.

present the confusion matrix for the prospective set and the combined samples, respectively.

DISCUSSION

The aim of the present study was to design algorithms with clinical history predictors for the diagnosis of BL allergy. If a reliable diagnosis could be achieved, then complete drug allergy evaluation would become unnecessary, at least for some subgroups of patients. Candidate predictors were obtained from patients' demographic characteristics and clinical history. We used the result of drug allergy evaluation, including ST and DPT, as a reference standard.

The prevalence of allergy in the 2 samples was different (23.6% on the retrospective vs 31% on the prospective sets, respectively), and for both samples it was slightly higher than our previous reports^{14,15} but the prevalence of BL allergy varies according to the study population and is subject to the inevitable selection bias of drug allergy cohorts completing drug allergy evaluation.¹⁶⁻¹⁸ Moreover, the prevalence of allergy might have slightly been underestimated in our population. Because of the need for a confirmed diagnosis for the comparison, we excluded severe life-threatening skin reactions from this study (severe cutaneous allergic reactions), where, owing to ethical reasons, the patient could not be challenged if ST result were negative.¹⁹ However, this latter group only comprised 9 patients in the retrospective sample (0.4%) and are unlikely to have influenced the results.

Based on the receiver operating characteristic curves, a model with a high sensitivity (>95%) would retain only 8.6% specificity and result in an overdiagnosis of BL allergy as is the case currently when diagnosis is based on clinical assessment only.

We identified some clinical risk factors for BL allergy through multivariate analysis. The highest OR was found for a clinical history of anaphylactic shock, which led to a 4-fold increase in the risk of being allergic. Other risk factors (anaphylaxis w/o shock, multiple episodes, and immediate reactions) had modest OR (2-fold increase in risk of BL allergy). Studies addressing drug allergy regularly search for risk factors.^{15,18,20} In the drug allergy field, except for a few specific cases associated with very high risks of developing severe DH,²¹ relative risk estimates (eg, OR) have little relevance to prognostication in practice. Few studies have tackled a predictive model approach. Dursun et al²² evaluated whether characteristics of patients and previous reactions could predict accurately the diagnosis of aspirin-exacerbated respiratory disease. Using logistic regression analysis, a model including age, the number of previous reactions, and a score for the sense of smell was designed. The model classified correctly 86% of all subjects and 99.5% of patients with aspirin-exacerbated respiratory disease. The high performance of the model is probably related to the study population (patients with chronic sinusitis, nasal polyps, and asthma) and the specific profile of aspirin hypersensitivity in this population. Mostly however, despite historical reactions that sound convincing, a substantial number of patients with BL allergy

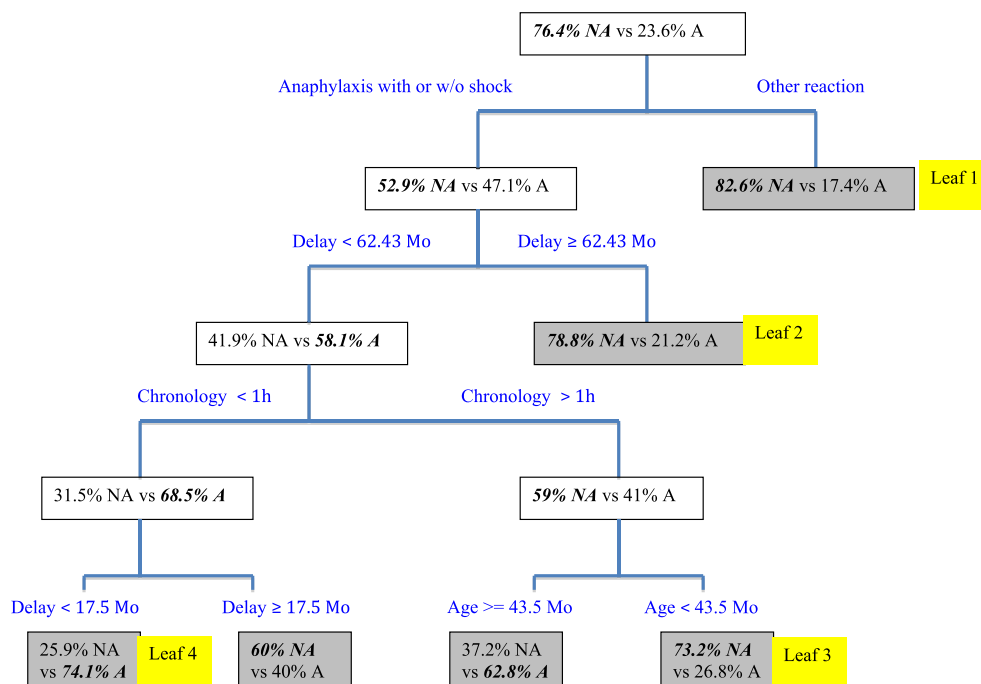


FIGURE 2. Decision tree designed on the retrospective sample. The gray nodes are terminal nodes (no more splitting afterward). The predicted values for each node (A, NA) are presented in italics (bold). A, Allergic; NA, nonallergic.

suspicion will turn out not to be allergic. In our study, logistic regression provided a model containing 9 clinical history predictors, or combination of predictors, including all the risk factors we identified for BL allergy (anaphylactic reactions, age, multiple reactions, immediate chronology, penicillins). Because of the relatively high number of variables in the equation, the model would probably be difficult to use in practice, but its main pitfall is its low sensitivity. The increase in sensitivity is accompanied by a decrease in specificity. For the optimal threshold of the score (value of the YI for which sensitivity and specificity are maximized), the best combined performances on the retrospective sample were sensitivity = 51% and specificity = 75%, meaning that for this cutoff, 1 in 2 allergic patients would be missed. This is unacceptable when exploring iatrogeny. For different cutoffs, the model achieved higher sensitivity and specificity on the validation sample. This sample included more patients with independent risk factors for BL allergy. However, we cannot exclude that higher performances may simply be due to chance.

Santurino et al²³ designed a retrospective model (using logistic regression) and a corresponding software to predict the diagnosis of DH including 8 variables in the medical history. In their series of 416 tested patients to several drugs (with 13.4% confirmed DH, 26.4% suspected DH, and 58.1% no DH), performances are slightly better than those of our logistic regression model, but the model suffers from the same pitfall: increasing specificity is accompanied by decreasing sensitivity (for a similar sensitivity with our study, the one that maximizes the YI, ie, sensitivity = 48%, specificity = 84%, PPV = 43%, NPV = 86%). No external validation was performed in this study and cases misclassified as nonallergic are not discussed, neither with respect to their clinical history nor according to their reaction during the allergy evaluation.

Regarding our decision tree, the misclassification rates on the retrospective and prospective samples were similar (19.4% vs 21%). Overall, 1 in 5 patients would be misclassified. The purpose of a decision tree is to produce groups of subjects that are as homogeneous as possible, with respect to the predicted variable ("pure leaves"). Sometimes, even if the leaves are pure, they have too few subjects, to be reliable, for future predictions. Our decision tree designed on the retrospective sample had 6 leaves, with at least 30 subjects/leaf, which is acceptable. Among these leaves, none was pure (no complete separation of allergic and nonallergic individuals). However, 3 leaves were particularly interesting, namely, the ones predicting nonallergic individuals with a high probability (three-fourth or more of the nonallergic individuals in leaves 1, 2, and 3). We subsequently analyzed the allergic patients who were misclassified in these leaves. Indeed, more than half of the allergic patients (70.5% on retrospective data and 56.4% on prospective data) were misclassified. Within the misclassified patients' group, the false-negative patients are the most problematic. On the retrospective sample, this group included 331 patients. Among them, 319 were distributed in leaves 1, 2, and 3, with 275, 26, and 18 patients, respectively. The 275 patients (83% of the total misclassified allergic patients) had described an episode of allergy with no severity criterion. It is known that clinicians tailor the drug allergy evaluation according to the severity of the initial reaction²⁴ although no standardized protocols exist. No special precaution would be demanded a priori for these patients, although they did turn out to be allergic. Within the misclassified allergic group, anaphylaxis (with or w/o shock) was elicited in 13 patients (3.1%) during allergy evaluation (ST or DPT). These patients were distributed in leaves 1, 2, and 3 and therefore do not constitute a clearly identifiable group. Therefore, although the decision tree seems to put forward some interesting

subgroups in leaves 1, 2, and 3, this information could hardly be extrapolated at the individual level to allow tailoring of the allergy evaluation without endangering a nonnegligible number of patients. Indeed, false-negative results have a medical cost but an ethical cost as well (reoccurrence of iatrogeny). As for the logistic model, the sensitivity of the decision tree was higher on the prospective sample (43.5% vs 29.5%).

One strength of our study is to have performed external validation (in a prospective sample), to avoid overfitting of the models on the training data. When compared, the 2 samples showed significant differences. Moreover, the prospective study was carried out in a specific time frame, possibly adding some selection bias. However, generally, the diagnostic accuracy of the 2 models on prospective data was similar to their performance on the retrospective data, which means that the models are rather stable and can probably barely be improved from a statistical point of view. It could be argued that a more thorough clinical history evaluation (ie, increasing the number of clinical parameters) could improve the performance of such a model. In the study of Santurino et al,²³ of 24 potential clinical history variables, only 8 were selected for the final model, and the performances were slightly higher, but still not satisfactory. The poor overall performance of the models is probably related to the poor predictive value of clinical history data, which is generally accepted.¹⁹ This emphasizes the need to search for other predictors¹⁷ and biological or genome data could be the answer to these unmet needs. In the study of Benahmed et al,²⁵ a pharmacovigilance algorithm²⁶ was used in 677 consecutive patients with 1001 reactions suggestive of DH (all drugs combined). This assessment of causality depends on criteria that are divided into 2 groups: chronology and symptoms or signs. Sensitivity of the algorithm was 10.3% (95% CI, 5.8%-14.8%) and specificity was 76.9% (95% CI, 74.0%-79.8%). When the “unknown scores” (20.4% of clinical histories could not be assigned a probability score because of a lack of some criteria required for scoring) were excluded, sensitivity and specificity were 12.0% (95% CI, 6.8%-17.2%) and 97.8% (95% CI, 96.7%-98.9%), respectively. In addition, when 1 criterion (ie, the definition of “reintroduction of the culprit drug”) of the Begaud algorithm was modified, the sensitivity became higher (31%; 95% CI, 23.5%-38.5%) for a similar specificity of 95% (95% CI, 93.3%-96.7%). Our 2 models show better sensitivity probably because the criteria take into account other independent risk factors for BL allergy, besides chronology and semiology. Whether current drug allergy biology could help in improving the models is unlikely, due to the low sensitivity of BL-specific IgE assays or the technically cumbersome lymphocyte transformation/activation assays.²⁷

In a recent report from Massachusetts General Hospital, Blumenthal et al²⁸ presented a straightforward clinical guideline for inpatients with a suspicion of allergy to penicillins. This guideline used clinical history and possible test dosing when appropriate in most cases without the necessity for ST. The implementation of this guideline led to a 7-fold increase in the use of BL test dosing without an increase in adverse reactions. Most patients were switched to third- and fourth-generation cephalosporins. This attitude is pragmatic and useful in the short term but the advantage of a negative drug allergy evaluation is that the false BL allergy diagnosis can be removed, whereas performing a challenge with cephalosporins does not rule out the penicillin allergy or the need to perform challenges with other cephalosporins in the future.²⁹

Doctors are bound to the principle of *primum non nocere* during their practice. As allergists, we are supposed to shed light on the utility of drug avoidance when a suspicion of DH occurs. Contrary to our peers in other areas in medicine, we test iatrogeny, to either confirm it or rule it out. Good clinical judgment, good tools, and good practice are essential to minimize the risks of iatrogeny reoccurrence, following medical decision. In this study, our objective was to design models for the diagnosis of BL allergy, from clinical history predictors. We have proven, on the basis of same samples and by means of 2 different independent statistical methods, that a combination of clinical history predictors cannot accurately differentiate between allergic and nonallergic individuals, even when different approaches are used. Nevertheless, some risk factors could be identified. Anaphylaxis remains an obvious independent clinical history marker, with the highest OR for confirming BL allergy, followed by a clinical history of immediate reactions and multiple episodes. Clinical history predictors are essential in organizing the allergy evaluation and in making sure that good practice is implemented during testing. This tool, as recommended at present by the learned societies, remains the most sensitive and specific one in proving or refuting iatrogeny related to DH.

Acknowledgements

We thank the investigators who participated in the patients' recruitment (Dr Christine Fabre, Dr Christian Gallen, and Dr Amparo Siret). We thank our collaborators from the Clinical Research Department for their assistance throughout the study, especially Laetitia Ferrando, our clinical research technician.

REFERENCES

1. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham 17 prognostic index in primary breast cancer. *Breast Cancer Res Treat* 1992;22:207-19.
2. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619-36.
3. Blanca M, Romano A, Torres MJ, Fernandez J, Mayorga C, Rodriguez J, et al. Update on the evaluation of hypersensitivity reactions to betalactams. *Allergy* 2009;64:183-93.
4. Macy E. Penicillin allergy: optimizing diagnostic protocols, public health implications, and future research needs. *Curr Opin Allergy Clin Immunol* 2015;15:308-13.
5. Bousquet PJ, Kvedariene V, Co-Minh HB, Martins P, Rongier M, Arnoux B, et al. Clinical presentation and time course in hypersensitivity reactions to betalactams. *Allergy* 2007;62:872-6.
6. Romano A, Blanca M, Torres MJ, Bircher A, Aberer W, Brockow K, et al. Diagnosis of nonimmediate reactions to beta-lactam antibiotics. *Allergy* 2004;59:1153-60.
7. Torres MJ, Blanca M, Fernandez J, Romano A, Weck A, Aberer W, et al. Diagnosis of immediate allergic reactions to beta-lactam antibiotics. *Allergy* 2003;58:961-72.
8. Demoly P, Kropf R, Bircher A, Pichler WJ. Drug hypersensitivity: questionnaire. EAACI Interest Group on Drug Hypersensitivity. *Allergy* 1999;54:999-1003.
9. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;338:b604.
10. España PP, Capelastegui A, Gorordo I, Esteban C, Oribe M, Ortega M, et al. Development and validation of a clinical prediction rule for severe community-acquired pneumonia. *Am J Respir Crit Care Med* 2006;174:1249-56.
11. Diepgen TL, Sauerbrei W, Fartasch M. Development and validation of diagnostic scores for atopic dermatitis incorporating criteria of data quality and practical usefulness. *J Clin Epidemiol* 1996;49:1031-8.
12. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
13. Breiman L, Friedman J, Olshen R, Stone C. Classification and regression trees. Belmont, Calif: Wadsworth; 1984.

14. Bousquet PJ, Pipet A, Bousquet-Rouanet L, Demoly P. Oral challenges are needed in the diagnosis of beta-lactam hypersensitivity. *Clin Exp Allergy* 2008; 38:185-90.
15. Rubio M, Bousquet PJ, Gomes E, Romano A, Demoly P. Results of drug hypersensitivity evaluations in a large group of children and adults. *Clin Exp Allergy* 2012;42:123-30.
16. Atanaskovic-Markovic M, Gaeta F, Medjo B, Gavrovic-Jankulovic M, Velickovic TC, Tmusic V, et al. Non-immediate hypersensitivity reactions to beta-lactam antibiotics in children—our 10 years experience in allergy work-up. *Pediatr Allergy Immunol* 2016;27:533-8.
17. Romano A, Gaeta F, Valluzzi RL, Caruso C, Rumi G, Bousquet PJ. IgE-mediated hypersensitivity to cephalosporins: cross-reactivity and tolerability of penicillins, monobactams, and carbapenems. *J Allergy Clin Immunol* 2010;126: 994-9.
18. Ponvert C, Perrin Y, Bados-Albiero A, Le Bourgeois M, Karila C, Delacourt C, et al. Allergy to betalactam antibiotics in children: results of a 20-year study based on clinical history, skin and challenge tests. *Pediatr Allergy Immunol* 2011;22:411-8.
19. Demoly P, Adkinson NF, Brockow K, Castells M, Chiriac AM, Greenberger PA, et al. International consensus on drug allergy. *Allergy* 2014; 69:420-37.
20. Romano A, Gaeta F, Valluzzi RL, Caruso C, Alonzi C, Viola M, et al. Diagnosing nonimmediate reactions to cephalosporins. *J Allergy Clin Immunol* 2012;129:1166-9.
21. Pavlos R, Mallal S, Phillips E. HLA and pharmacogenetics of drug hypersensitivity. *Pharmacogenomics* 2012;13:1285-306.
22. Dursun AB, Woessner KA, Simon RA, Karasoy D, Stevenson DD. Predicting outcomes of oral aspirin challenges in patients with asthma, nasal polyps, and chronic sinusitis. *Ann Allergy Asthma Immunol* 2008;100:420-5.
23. Hierro Santurino B, Mateos Conde J, Cabero Morán MT, Mirón Canelo JA, Armentia Medina A. A predictive model for the diagnosis of allergic drug reactions according to the medical history. *J Allergy Clin Immunol Pract* 2016;4: 292-300.
24. Co Minh HB, Bousquet PJ, Fontaine C, Kvedariene V, Demoly P. Systemic reactions during skin tests with beta-lactams: a risk factor analysis. *J Allergy Clin Immunol* 2006;117:466-8.
25. Benahmed S, Picot MC, Dumas F, Demoly P. Accuracy of a pharmacovigilance algorithm in diagnosing drug hypersensitivity reactions. *Arch Intern Med* 2005; 165:1500-5.
26. Begaud B, Evreux JC, Jouglard J, Lagier G. Imputabilité des effets inattendus ou toxiques des médicaments: actualisation de la méthode utilisée en France. *Thérapie* 1985;40:111-8.
27. Mayorga C, Celik G, Rouzair P, Whitaker P, Bonadonna P, Cernadas JR, et al. In vitro tests for Drug Allergy Task Force of EAACI Drug Interest Group. In vitro tests for drug hypersensitivity reactions. An ENDA/EAACI Drug Allergy Interest Group Position Paper. *Allergy* 2016;71:1103-34.
28. Blumenthal KG, Shenoy ES, Varughese CA, Hurwitz S, Hooper DC, Banerji A. Impact of a clinical guideline for prescribing antibiotics to inpatients reporting penicillin or cephalosporin allergy. *Ann Allergy Asthma Immunol* 2015;115: 294-300.
29. Solensky R. A novel approach to improving antibiotic selection in patients reporting penicillin allergy. *Ann Allergy Asthma Immunol* 2015;115:257-8.

ONLINE REPOSITORY

```
> print (tree)
```

```
n= 1991
```

```
node), split, n, loss, yval, (yprob)
```

```
1) root 1991 470 NonAllergic (0.2360623 0.7639377)
  2) semiology=anaphylaxis, anaphylactic shock 414 195 NonAllergic (0.4710145
0.5289855)
    4) delay< 62.43333 291 122 Allergic (0.5807560 0.4192440)
      8) chronology=< 1h 181 57 Allergic (0.6850829 0.3149171)
        16) delay< 17.56667 151 39 Allergic (0.7417219 0.2582781) *
        17) delay>=17.56667 30 12 NonAllergic (0.4000000 0.6000000) *
      9) chronology=> 1h 110 45 NonAllergic (0.4090909 0.5909091)
        18) age>=43.5 43 16 Allergic (0.6279070 0.3720930) *
        19) age< 43.5 67 18 NonAllergic (0.2686567 0.7313433) *
    5) delay>=62.43333 123 26 NonAllergic (0.2113821 0.7886179) *
  3) semiology=other than anaphylaxis and anaphylactic shock 1577 275 NonAllergic
(0.1743817 0.8256183) *
```

¹ node) = number of the node
split= variable used for discrimination
n= number of observations (patients)/node
loss= number of misclassified observations
yval= predicted value (allergic, All, or non allergic, NonAll)
yprob= probabilities of the predicted values
* terminal nodes = leaves

FIGURE E1. Printout of the decision tree designed on the retrospective sample. *loss*, Number of misclassified observations; *n*, number of observations (patients)/node; *node*, number of the node; *split*, variable used for discrimination; *yprob*, probabilities of the predicted values; *yval*, predicted value (allergic, All, or nonallergic, NonAll). *Terminal nodes = leaves.

APPENDIX E1. Patient collected data

- sex
- atopy*
- age at the reaction
- asthma†
- clinical presentation of the reaction
 - urticaria/angioedema/pruritus
 - anaphylaxis‡ w/o shock
 - anaphylactic shock
 - maculopapular exanthema
 - nonspecific cutaneous symptoms
 - isolated respiratory symptoms (bronchial hyperresponsiveness, rhinoconjunctivitis, other nonspecified respiratory symptoms)
 - other nonspecific or unknown symptoms
 - severe delayed reactions including SCARs§ and cytopenia
- chronology of the reaction||
- number of clinical histories related to BL/patient¶
- BL involved (penicillins, cephalosporins, other BLs)
- delay between the reaction and the allergy workup.

AGEP, Acute generalized exanthematous pustulosis; DRESS, Drug Reaction with Eosinophilia and Systemic Symptoms; GINA, Global Initiative for Asthma; SCARs, severe cutaneous allergic reactions; SJS, Stevens-Johnson syndrome.

At the end of the descriptive analysis, after observing the size of the different categories of the qualitative predictors, the following clinically meaningful groups were established:

- antibiotic: “penicillins” and “all other BLs”;
- chronology: “<1 h” (immediate) and “≥1 h” (nonimmediate);
- semiology: “anaphylactic shock,” “anaphylaxis w/o shock,” and “other reactions”; and
- number of reactions/patient: “1 episode” or “multiple episodes.”

The linearity assumption was verified for the variable “delay” and, therefore, this predictor was introduced as a continuous variable. The variable “age” was turned into dichotomy and was introduced in the analysis as classes, <18 y (children) and ≥18 y (adults) at the time of the reaction. Drug allergy is known to be more prevalent in adults than in children, and, therefore, classifying patients into groups of age was clinically pertinent.

*Defined as a positive ST to at least 1 of the common respiratory allergens in the Southern France area.

†Defined according to GINA 2014.^{E1}

‡Defined according to Sampson et al.^{E2}

§Severe cutaneous allergic reactions, SCARs (DRESS, AGEP, SJS/Lyell, vasculitis).

||Defined as the delay between the last drug intake and the occurrence of symptoms. Several delays were considered initially: (<1 h), (1-6 h), (6-24 h), (>24 h), and not determined (ND).

¶One episode or multiple episodes (this latter category was divided into single reactors, ie, patients having multiple episodes to the same BL, and multiple reactors, ie, patients having multiple episodes to different BLs).

TABLE E1. Confusion matrix for the training set

Diagnosis	Predicted by Tree	
	Allergic	Nonallergic
Actual diagnosis		
Allergic	139	331
Nonallergic	55	1466

Each column of the confusion matrix represents the categories (allergic/nonallergic) in the predicted class (ie, predicted by tree), and each row represents the categories in the actual class (ie, actual diagnosis of BL allergy following BL evaluation).

TABLE E2. Confusion matrix for the validation set

Diagnosis	Predicted by Tree	
	Allergic	Nonallergic
Actual diagnosis		
Allergic	27	35
Nonallergic	7	131

Each column of the confusion matrix represents the categories (allergic/nonallergic) in the predicted class (ie, predicted by tree), and each row represents the categories in the actual class (ie, actual diagnosis of BL allergy following BL evaluation).

TABLE E3. Confusion matrix for the whole sample (retrospective and prospective one, combined, for a total of 2191 patients)

Diagnosis	Predicted by Tree	
	Allergic	Nonallergic
Actual diagnosis		
Allergic	166	366
Nonallergic	62	1597

Each column of the confusion matrix represents the categories (allergic/nonallergic) in the predicted class (ie, predicted by tree), and each row represents the categories in the actual class (ie, actual diagnosis of BL allergy following BL evaluation).

REFERENCES

- E1. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention, 2015. Available from: www.ginasthma.org. Accessed June 29, 2017.
- E2. Sampson HA, Muñoz-Furlong A, Campbell RL, Adkinson NF Jr, Bock SA, Branum A, et al. Second symposium on the definition and management of anaphylaxis: summary report—Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. *J Allergy Clin Immunol* 2006;117:391-7.