European Journal of Epidemiology **17**: 19–29, 2001.

Classification methods for the identification of 'case' in epidemiological diagnosis of asthma

Mario Grassi, Simona Villani & Alessandra Marinoni for ECRHS Group
Faculty of Medicine, Health Science Department, University of Pavia, Pavia, Italy

Accepted in revised form 14 March 2001

Abstract. The identification of the asthmatic 'case' in epidemiological research is a controversial issue. This study was aimed at classifying asthmatic subjects using a statistical decision rule that minimised the misclassification rate with respect to the clinicians' diagnosis. The rule was defined by a combination of predictors that are easily observed in epidemiological studies (asthma-like questions, physiological tests) without necessarily including the clinical opinion of expert physicians. From pooled data on 1103 subjects at the three Italian centres of the European Community Respiratory Health Survey (ECRHS) a post-consensus clinicians' diagnosis of asthma was obtained, and seven predictors were selected from among 18 potential candidates (specificity ranged from 64 to 99%, but sensitivity ranged from 22 to 62%). This data set was processed with tree-structured classifier techniques (the Classification And Regression Trees, CART), classical discriminant analysis (Fisher's Linear Discriminant Function, LDF), and the neural network method (Multi-Layer Perceptron, MLP model). The results suggest that modifications of the 'classification tree' provide a more useful decision rule, sensitive (93%) and specific (85%), than either LDF or MLP. The decision tree is readily interpretable from a clinical perspective and uses five out of the seven predictors (in descending hierarchical order: ever had asthma, current asthma, shortness of breath, atopy and wheezing and breathless). The findings seem to indicate a considerable success with respect to previous epidemiological studies and await repetition in other ECHRS populations.

Key words: Asthma, Classification tree, Clinicians' diagnosis, European Community Respiratory Health Survey

Introduction

Asthma is a serious international health problem for all patients of all ages. There was a dramatic increase in asthma mortality throughout the 1970s and 1980s in many industrialised countries, including US, Japan, New Zealand, England and Wales [1]. The worrying trend in asthma mortality has drawn the attention of researchers, who classified some of these deaths as potentially preventable (avoidable mortality).

In order to prevent this disease, it is fundamental to have a general definition (all-encompassing) of asthma, which is still a debated issue today, as well as some criteria for recognising and diagnosing asthma [1]. In the past asthma has been defined in the 1959 Ciba Foundation Symposium [2] as '...the condition of subjects with widespread narrowing of the bronchial airways, which changes its severity over short periods of time either spontaneously or under treatment…', by the American Thoracic Society in 1962 [3] as '...a disease characterised by an increased responsiveness of the tracheae and bronchi to various stimuli and manifested by a widespread narrowing of the airways that changes in severity either spontaneously or as a result of therapy' and by WHO in 1975 [4] as '...a chronic condition characterised by recurrent bronchospasm resulting from a tendency to develop reversible narrowing of the airway lumina in response to stimuli of a level or intensity not inducing such narrowing in most individuals'. All these definitions have proved to be unsatisfactory in one way or another and according to the perspective from which clinicians and researchers view the problem [5].

In epidemiological studies before 1980 asthma was defined on the basis of responses to a questionnaire regarding the presence or absence of particular respiratory symptoms generally associated with asthma. After recognising that questionnaire responses could be influenced by a wide variety of cultural, sociological and psychological factors, researchers developed a second definition. This implied positivity to a bronchial hyperresponsiveness (BHR) test. Soon it became apparent through clinical and population studies that the BHR was neither sensitive nor specific as a test for asthma. In 1992 Toelle et al. [6] proposed the incorporation of two previous definitions to use as an epidemiological definition of asthma; they defined *current asthma* as the presence of symptomatic BHR, that is hyperresponsiveness to a challenge test, plus a positive questionnaire reply to recent wheeze or exercise wheeze.

Recently, an expert panel from different countries formulated, on the basis of past definitions and with the addition of specific cellular functions, the following definition of asthma: '... a respiratory disorder (observed as the presence of particular symptoms) characterised by variable airflow obstruction to which specific alterations of cellular functions are accompanied; breathlessness and whistling, cough, dyspnoea, chest tightness, reversibility to bronchodilators and corticosteroids, increased airway responsiveness to a variety of stimuli evidence of inflammation in which eosinophils, mast cells, lymphocytes together with a multitude of cytokines have important roles' [7]. Nevertheless this definition cannot be fully applied to epidemiological research, due to absence of an objective and non-invasive measurement of airflow limitation.

Burr, in the editorial that appeared in Allergy in 1992 [8], affirms that 'since asthma is a clinical diagnosis, it would seem reasonable to use a respiratory physician (or a group of physicians) as arbiter of what is and what is not asthma'. The Italian group of the European Community Respiratory Health Survey (ECRHS) has put forth a similar proposal [9]: to classify cases of asthma by agreement in clinical judgement formulated, separately, by three experts on the basis of a standardised clinical interview with a questionnaire, respiratory function tests and allergological tests. An agreement between expert panels generally yields good results [10, 11], so that a post consensus classification of asthma seems to be a highly reproducible and stable method for identifying 'non-asthmatic' and 'asthmatic' subjects. The Italian researches suggest that the post consensus diagnosis could be a 'gold standard' for evaluating the sensitivity and specificity of epidemiological tools. Similarly, Jenkins et al. [12] used the physician assessment of current asthma as a gold standard among the available measures. They determined the validity of self-reported asthma (questionnaire) by agreement with physician diagnosis of asthma.

However, it is often difficult to formulate a clinical judgement about symptoms and clinical tests, while it is easier to obtain the last two types of information. It would be useful to be able to identify the asthmatic status of a subject on the basis of only the clinical test results and the symptoms manifested (predictors of asthmatic or non-asthmatic group).

Aim

The aim of the present paper is: to identify asthmatic subjects using a statistical decision rule, that is readily interpretable from a clinical perspective, minimising the misclassification rate with respect to a clinician's diagnosis, and defined by a sensitive combination of predictors (asthma-like questions, self-consciousness of asthmatic disease, physiological tests) readily observed in epidemiological studies, without necessarily including the clinical opinion of expert physicians.

Materials and methods

In the present study we analysed data from the Italian ECRHS centres. The design of the ECRHS has been already described elsewhere [13, 14]. Briefly, the survey consisted of two stages. In the first one, a screening questionnaire with 10 questions was mailed to a random sample of 20–44 year old (M:F = 1:1) living in three Italian areas. In the second stage, a random sample of subjects who responded to the questionnaire, plus symptomatic subjects from stage I (who had reported asthma attacks, use of anti-asthmatic drugs, or awakening due to shortness of breath), were submitted to other checks by clinicians (a standardised clinical interview as well as lung function and allergy tests).

The sample and the variables

The sample examined here corresponds to the ECRHS stage II Italian sample [15–19]. This was composed of 1104 subjects: 376 from Pavia, 355 from Turin and 373 from Verona. Information (483 variables) was collected for each subject by means of screening questionnaire, standardised clinical interviews, respiratory function tests, allergy tests, and personal characteristics. For the purposes of the present paper, we examined the 19 variables listed in Table 1. One subject was eliminated from the sample because all data regarding to the selected variables was lacking while the information was complete for 811 subjects. A missing at random pattern was observed for the other subjects with very few incomplete data (with respect the 19 variables examined). Therefore, we used the sample mode (or mean) of a binary (or numerical) variable to replace any missing data for that variable.

The full protocol that was followed to define each subject as 'asthmatic' or 'not asthmatic' (post-consensus clinicians' diagnosis of asthma) is described elsewhere [9]. Briefly, in each centre subjects were classified according to the diagnosis formulated (independently) by three experienced clinicians (pneumologists and/or allergologists). Diagnosis was initially based on the responses to a standardised clinical interview; secondly on the lung function and methacholine challenge test, and, finally, immunoglobulin E (IgE) and skin-prick tests. The introduction of a new source of information had to confirm or modify the diagnosis formulated until previous step. Each physician formulated a diagnosis using his own experience and knowledge, without consulting the other two physicians. Diagnosis was immediately formulated if there was agreement among the clinicians. On the other hand in the case of a disagree-

Table 1. Definition and coding of the variables used as potential predictors (V2–V19) of response variable 'post-consensus clinicians' diagnosis' of asthma (V1)

	Variables	Definition and coding
V1	Post-consensus clinicians' diagnosis of asthma	Non-asthmatic = 1 Asthmatic = 2
V2	Age	Given by the difference between the date of birth of the subject and the date of the interview. Range = 20–44; mean ± SD = 33.6 ± 6.8
V3	Sex	Male = 1 Female = 2
V4	Wheeze and breathless	*Being breathless when the wheezing noise was present (in the last 12 months)* No = 1 Yes = 2
V5	Wheeze without a cold	*Wheezing or whistling in the chest in the absence of a cold (in the last 12 months)* No = 1 Yes = 2
V6	Chest tightness	*Waking up with a feeling of tightness in the chest (in the last 12 months)* No = 1 Yes = 2
V7	Breathless at rest	*Having an unexpected attack of breathlessness at rest (in the last 12 months)* No = 1 Yes = 2
V8	Breathless after effort	*Having an unexpected attack of breathlessness after intense effort (in the last 12 months)* No = 1 Yes = 2
V9	Shortness of breath	*Being awakened by an attack of shortness of breath at any time (in the last 12 months)* No = 1 Yes = 2
V10	Waking with cough	*Being awakened by an attack of coughing at any time (in the last 12 months)* No = 1 Yes = 2
V11	Ever asthma	*Have you ever had asthma?* No = 1 Yes = 2
V12	Attack of asthma	*Having had an attack of asthma at any time (in the last 12 months)* No = 1 Yes = 2
V13	Asthma by doctor	*Asthma confirmed by a doctor* No = 1 Yes = 2
V14	Asthma medicine	*Currently taking any medicines (including inhalers, aerosols or tablets) for asthma* No = 1 Yes = 2
V15	Atopy	Negative S*kin* P*rick* T*est* to all of a panel* of allergens = 1 Positive S*kin* P*rick* T*est* to any one of a panel* of allergens = 2 *Tested allergens: E1 cat, *Dermatophagoides pteronyssinus*, M2 *Cladosporium herbarum*, M6 *Alternaria tenuis*, G6 Timothy grass, T9 Olives, T3 Birch, W21 *Parietaria judaica*, *Artemisia vulgaris* (local allergens 1), *Dermatophagoides farinae* (local allergens 2), W1 common Ragweed

Table 1. (*Continued*)

	Variables	Definition and coding
V16	Pulmonary injury	No impairment = FEV1 $\geqslant$ 80% *and* FVC $\geqslant$ 80% *and* FEV1/FVC $\geqslant$ 75% = 1 Mild impairment = FEV1 = 60–79% *or* FVC = 60–79% *or* FEV1/FVC = 60–74% = 2 Moderate impairment = FEV1 = 41–59% *or* FVC = 51–59% *or* FEV1/FVC = 41–59% = 3 Severe impairment = FEV1 $\leqslant$ 40% *or* FVC $\leqslant$ 50% *or* FEV1/FVC $\leqslant$ 40% = 4 *FEV1 = Forced Expiratory Volume in 1 sec; FVC = Forced expiratory Vital Capacity
V17	BHR	Subjects with an initial FEV1 less than 70% predicted or a PD_{20} less than 2 mg methacholine are defined as having bronchial hyperresponsiveness No = 1 Yes = 2
V18	Family history	Given by the sum of 8 binary items* (the greater the value, the greater the family history for asthma....) **How many of your brothers ever had asthma?; How many of your brothers ever had eczema, skin or nasal allergy or hay fever?; How many of your sisters ever had asthma?; How many of your sisters ever had eczema, skin or nasal allergy or hay fever?; Has your mother had asthma?; Has your mother ever had eczema, skin or nasal allergy or hay fever?; Has your father had asthma?; Has your father ever had eczema, skin or nasal allergy or hay fever?*
V19	Current asthma	Subjects with BHR at testing and a positive questionnaire reply to V4 or V5 or V8 are defined as having current asthma No = 1 Yes = 2

ment, the diagnosis was formulated when the experts reached major agreement, after discussing all the data. Agreement among clinicians within each centre was good (Cohen κ coefficient was 0.71) as well as the agreement in the consensus diagnosis made by the panel of experts between centre (Cohen κ coefficient was 0.88) [9, 19]. In this paper, the variable 'post-consensus clinicians' diagnosis of asthma' denoted the response variable, and the predictor variables were the following (Table 1): two background variables (age and sex); eleven questions subdivided in past-year symptoms (V4–V10) and self-reported asthma (V11–V14), taken from the clinical standardised interview; three clinical variables (atopy, pulmonary injury and BHR) defined by the results of physiological tests, according to criteria widely reported in the literature [13, 20, 21]; family history, obtained from a combination of questions. Finally, among the predictors, we considered the variable current asthma, which was constructed taking into consideration symptoms and BHR, as suggested by Toelle et al. [6] in their epidemiological definition of asthma. In particular, in the present paper a subject was defined as having 'current asthma' if he showed hyperresponsiveness to a challenge test and, if during the last 12 months presented breathlessness with wheezing noise, wheezing or whistling in the chest in the absence of a cold, or had an unexpected attack of breathlessness after intense effort.

Statistical analysis

In addition to usual univariate statistical analysis, given the large number of predictor variables for the response variable 'post-consensus clinicians' diagnosis of asthma', we performed a preliminary selection of the variables using an 'all subset discriminant analysis' procedure [22, 23]. On the data set selected from the previous exploratory analysis, some classification techniques were attempted, including the tree-structured classifier technique (the Classification And Regression Trees, CART), classical discriminant analysis (Fisher's Linear Discriminant Function, LDF), and neural network method (Multi-Layer Perceptron, MLP model), with the goal of developing decision rules for allocating a subject either to the non-case (non-asthmatic) or the case (asthmatic) group.

A recent, authoritative review of the statistical methodology for CART, LDF and MLP is described by Hand [22]. Briefly, CART constructs a classification tree using a sequence of binary splits for the predictor variables on which to discriminate the two groups (case or non-cases). The purpose is to reveal the subdivision, or partitions, of the data set with respect to distinct combination of predictor variables that jointly influence group status. LDF attempts to design a decision rule by forming linear combinations of the predictor variables, $f(x) = a_0 + a_1x_1 + a_2x_2 + \cdots + a_px_p$ which best discriminates between the two

Table 2. Overall prevalence (Pv), specificity (Sp), sensitivity (Se), accuracy (Ac), positive predictive value (PPV) and negative predictive value (NPV) of the potential predictors of 'post-consensus clinicians diagnosis of asthma' in the study sample (n = 1103)

	Variables	Pv (%)	Sp (%)	Se (%)	Ac (%)	PPV (%)	NPV (%)
V4	**Wheeze and breathless**	**4.5**	**99.04**	**25.47**	**88.30**	**82.00**	**88.60**
V5	Wheeze without a cold	12.0	93.31	42.86	85.95	52.27	90.53
V6	Chest tightness	11.5	93.31	39.75	85.49	50.39	90.06
V7	Breathless at rest	8.4	95.22	29.81	85.68	51.61	88.81
V8	**Breathless after effort**	**13.7**	**90.76**	**39.75**	**83.32**	**42.38**	**89.81**
V9	**Shortness of breath**	**11.2**	**93.42**	**38.51**	**85.40**	**50.00**	**89.89**
V10	Waking with cough	28.9	74.10	46.58	70.08	23.51	89.03
V11	**Ever asthma**	**11.4**	**97.24**	**62.11**	**92.11**	**79.37**	**93.76**
V12	Attack of asthma	5.0	99.79	32.92	90.03	96.36	89.69
V13	Asthma by doctor	10.5	97.24	55.90	91.21	77.59	92.81
V14	Asthma medicine	3.3	99.68	20.50	88.12	91.67	88.00
V15	**Atopy**	**22.2**	**83.65**	**56.52**	**79.69**	**37.14**	**91.84**
V16	Pulmonary injury	11.5	90.23	21.74	80.24	27.56	87.09
V17	**BHR**	23.5	**81.21**	**50.93**	**76.79**	**31.66**	**90.64**
V18	Family history	39.5	63.38	56.52	62.38	20.87	89.51
V19	**Current asthma**	**7.0**	**97.72**	**34.16**	**88.40**	**71.43**	**89.67**

The predictors selected by 'all subset discriminant analysis' are printed in boldface.

groups. The optimisation problem then becomes one of estimating coefficients of the predictors variables so that we classify the subjects as case if $f(x) > 0$ or as non-case if $f(x) < 0$. MLP connects a successive linking of predictor variables, each corresponding to one 'input neurons', and the group status, each corresponding to 'output neurons', using hidden layers with a number of neurones. For a given network architecture and input set, the output is entirely determined by the weights and by the non-linear function (activation) that link outputs to the inputs. Optimal weights were determined to have better predictive output accuracy. We consider a MLP model with three layers (input, hidden, output).

We used SPAD 3.01 statistical package Windows 3.x/95 version [23], which allows the three procedures to be performed automatically within the same framework. The criteria of decision rules we have chosen are reported in the Appendix.

Results

The sample was made up of 50.3% females, mean age 33.6 years (±6.8 years). The prevalence of respiratory symptoms (Table 2) was ranged between 4.5% (wheeze and breathless) and 29% (in waking with cough). 11.4% of the subjects said they had asthma in the past, 10.5% declared they had had asthma confirmed by a family doctor, 5% reported an asthma attack in the last 12 months and only 3.3% said they were currently taking therapy for asthma. Furthermore, 23.5% were classified as having BHR and 11.5% with pulmonary injury. Elevated percentages of subjects positive to any one of a panel of allergens (atopy) and with at least one first degree relative with asthma, or eczema, or nasal allergy, or hay fever (family history) were found. Atopy was equal to 22.2% and family history to 39.5%. Using the 'post-consensus clinicians' diagnosis of asthma', the prevalence of asthmatic subjects was 14.5%, while according to the epidemiological definition of asthma, 'current asthma' was 7%. Table 2 reports prevalence for each predictor variable further than sensitivity, specificity and predictive values respect to response variable (post-consensus clinicians' diagnosis of asthma). The accuracy of past-symptoms variables (from V4 to V10) was, on average, 85%. Self-reported asthma variables plus 'current asthma' (from V11 to V14 and V19) had the highest (about 90%) accuracy. Clinical variables (from V15 to V18) had an accuracy of a little less than 80%. Finally, the accuracy of the family history variable was the lowest, being only 62%. The sensitivity was rather small for all predictors.

The predictors of the response variable, selected for classification analysis (CART, LDF and MLP) are listed in boldface in Table 2: V4 = wheezing and breathlessness; V8 = breathless after effort; V9 = shortness of breath; V11 = ever asthma; V15 = atopy; V17 = BHR; V19 = current asthma. The selection of predictors is the result of an all subset discriminant analysis procedure. In particular, the variables of the self-reported asthma subset are highly correlated, and so only 'ever asthma' was selected, while three out of four (V4, V8 and V17) of the single components of 'current asthma' and the same 'current asthma' combination was selected as possible potential predictors.

Table 3 shows the results of three classification rules A, B and C (see Appendix for a more detailed discussion of decision criteria), obtained with CART

Table 3. 2×2 confusion tables obtained by CART analysis on seven predictor variables using three decision rules, having different options concerning prior probabilities (π_1 and π_2) and penalties (c_1 and c_2) and number (t) of nodes in the tree. Rule A: $\pi_1 = p_1$, $\pi_2 = p_2$, $c_1 = 1$, $c_2 = 2$ and $t = 4$; rule B: $\pi_1 = p_1$, $\pi_2 = p_2$, $c_1 = 1$, $c_2 = 10$ and $t = 7$; rule C: $\pi_1 = ½$, $\pi_2 = ½$, $c_1 = 1$, $c_2 = 1$ and $t = 7$. These 2×2 tables are obtained by 10-fold cross-validations

		Rule A		Rule B		Rule C	
		1	2	1	2	1	2
Post consensus	1	893	49	824	118	855	87
	2	40	121	20	141	27	134
		Specificity = 94.8%		Specificity = 87.5%		Specificity = 90.8%	
		Sensitivity = 75.2%		Sensitivity = 87.6%		Sensitivity = 83.2%	
		Accuracy = 91.9%		Accuracy = 87.5%		Accuracy = 89.7%	

analysis. Ten-fold cross-validations for the pruned subtree sequence suggested that the trees with $t = 4$, 7 and 7 nodes would be optimal for rules A, B and C respectively. The minmax error rate criteria are supported by rule B and minmax number criteria by rule A. We chose rule B because it had greater sensitivity than rule A and C.

The classification tree for post-consensus clinicians' diagnosis of asthma resulting from recursive partitioning analysis (rule B) is shown in Figure 1.

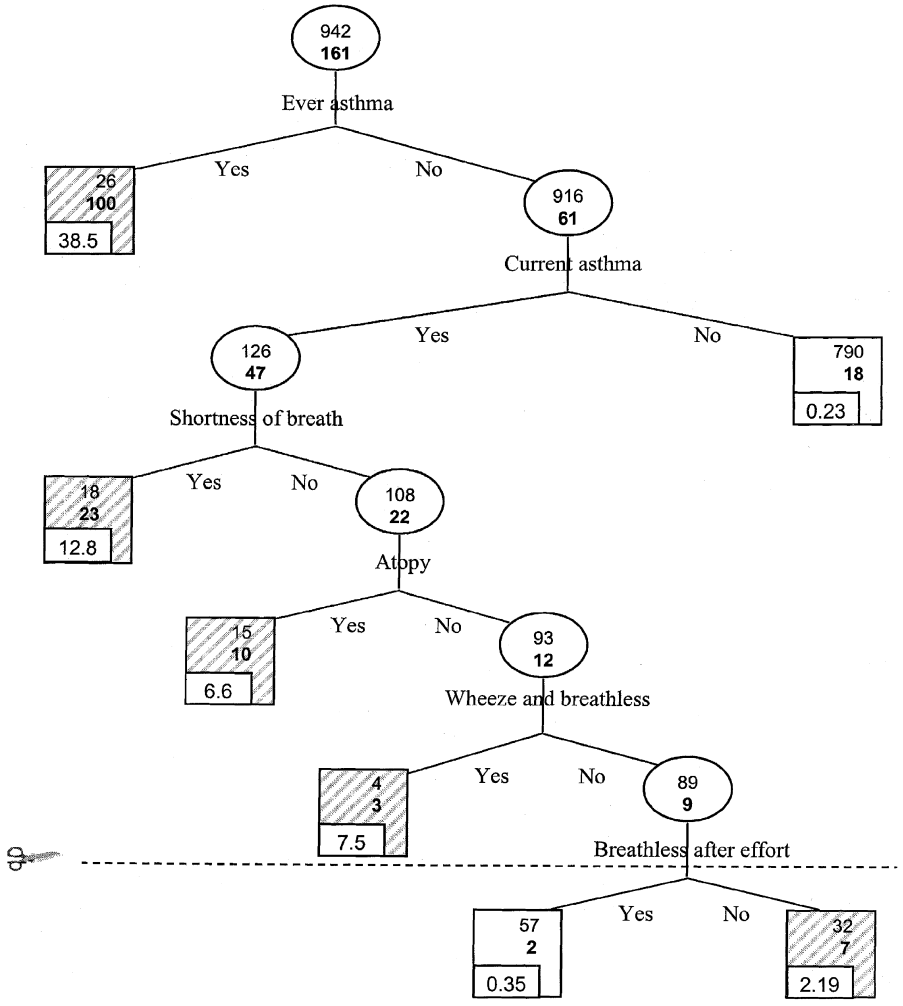

Figure 1. Classification tree for post consensus diagnosis of asthma in the study sample ($n = 1103$) with *ever asthma*, *current asthma*, *shortness of breath*, *atopy*, *wheeze and breathless* and *breathless after effort* as predictors. In each node (circles and squares) the boldface number (n_2) indicates the number of subjects at that node with a post consensus diagnosis of 'asthmatic' and the normal face number (n_1) indicates the subjects with post consensus diagnosis of 'non asthmatic'. The 'misclassification risk ratio' ($= n_2c_2/n_1c_1$) is reported in the small box in the lower left corner of the terminal nodes (squares): if the 'misclassification risk ratio' is greater than 1 the square is shaded.

For each intermediate node (circle) of the tree a question is asked and subjects whose answer is 'positive' are assigned to the left branch, and the others are assigned to the right branch. The squares are the terminal nodes, or 'leaves', of the tree; if the leaf is shaded (not shaded) the subjects are classified as 'asthmatic' ('non-asthmatic'). In rule B, the prior probabilities are fixed to the observed prevalences and 10 penalty points are charged for misclassifying subjects in the true 'asthmatic' class and one penalty point is charged for misclassifying subjects in the true 'non-asthmatic' class. The predicted class (1 = 'non-asthmatic' or 2 = 'asthmatic') results defined by the 'misclassification risk ratio', which is the ratio between the average cost error of allocating the 'leaf' to the class 1 = 'non-asthmatic' and the average cost error of allocating the 'leaf' to class 2 = 'asthmatic'. If the risk ratio is greater than one, the end node ('leaf') is allocated to 'asthmatic' class and vice versa. Furthermore, there are seven terminal nodes: five predicted 'asthmatic' classes and two 'not asthmatic' classes.

In each circle (or square) the boldface number indicates the number of subjects at that node with a post-consensus diagnosis of 'asthmatic', and the normal face number indicates the number of subjects with a post-consensus diagnosis of 'non-asthmatic'.

The seven node tree makes use of six variables (in descending order): V11 = ever asthma; V19 = current asthma; V9 = shortness of breath; V15 = atopy; V4 = wheezing and breathless; V8 = breathless after effort. We can summarise the tree as follows. The top node ('root') is split on the 'ever asthma' question. A positive answer indicates a post-consensus diagnosis 'asthmatic', while a negative answer is split on the 'current asthma' question. A negative answer indicates a post-consensus diagnosis of 'non-asthmatic', while a positive answer is divided on the 'shortness of breath' question, and so on. This tree correctly classifies 87.5% of the training data, using 10-fold cross-validation; the sensitivity which is equal to the specificity, gives a false positive error rate equal to the false negative error rate of 12.5%.

Detailed classification results for the other two methods, LDF and MLP, are given in Tables 4 and 5, respectively. Each table shows the results of the three classification rules A, B and C (see Appendix for a more detailed discussion of decision criteria), subsequently validated by the 10-fold cross-validation procedure as in CART.

The LDF rules have a better performance than those of CART and MLP do. The LDF rules correctly classify 92.7%, 87.5% and 93.1% of the training data for rules A, B and C, respectively. The MLP A, B and C rules show poor sensitivity, although the overall misclassification error rate is less than 8%. Moreover, we prefer the CART rule B as the better decision rule, even though LDF has a slight

Table 4. 2×2 confusion tables obtained by LDF analysis on seven predictor variables using three decision rules, with different options concerning prior probabilities (π_1 and π_2) and penalties (c_1 and c_2). Rule A: $\pi_1 = p_1$, $\pi_2 = p_2$, $c_1 = 0.09$, $c_2 = 0.91$; rule B: $\pi_1 = \pi_2 = ½$, $c_1 = 0.09$, $c_2 = 0.91$; rule C: $\pi_1 = \pi_2 = ½$, $c_1 = c_2 = 1$. These 2×2 tables are obtained by 10-fold cross-validations

		Rule A		Rule B		Rule C	
		1	2	1	2	1	2
Post consensus	1	892	50	821	121	907	35
	2	31	130	17	144	41	120
		Specificity = 94.7%		Specificity = 87.2%		Specificity = 96.4%	
		Sensitivity = 80.7%		Sensitivity = 89.4%		Sensitivity = 74.3%	
		Accuracy = 92.7%		Accuracy = 87.5%		Accuracy = 93.1%	

Table 5. 2×2 confusion tables obtained by MLP analysis on seven predictor variables using three decision rules, with different options concerning prior probabilities (π_1 and π_2), penalties (c_1 and c_2) and number (h) of neurones in one hidden layer. Rule A: $\pi_1 = \pi_2 = ½$, $c_1 = c_2 = 1$ and $h = 7$; rule B: $\pi_1 = \pi_2 = ½$, $c_1 = c_2 = 1$ and $h = 2$; rule C: $\pi_1 = 0.95$, $\pi_2 = 0.05$, $c_1 = c_2 = 1$ and $h = 14$. These 2×2 tables are obtained by 10-fold cross-validations

		Rule A		Rule B		Rule C	
		1	2	1	2	1	2
Post consensus	1	911	31	920	22	920	22
	2	51	110	61	100	64	97
		Specificity = 96.7%		Specificity = 97.7%		Specificity = 97.7%	
		Sensitivity = 68.3%		Sensitivity = 62.1%		Sensitivity = 60.2%	
		Accuracy = 92.6%		Accuracy = 92.5%		Accuracy = 93.2%	

advantage over CART in both sensitivity and specificity because the subgroups represented by the branches of the classification tree are more readily interpretable from a clinical perspective than discriminant coefficients.

Discussion

The recursive partitioning technique (CART) and linear discriminant analysis seems to sensitively, specifically identify asthmatic subjects using easily observable predictors in an epidemiological survey, while the same is not true for the non-linear MLP method. Nevertheless, the CART analysis happens to be most satisfactory methodological approach which constructs a classification tree of the logical process involved in formulating a diagnosis that is best suited to practical implementation in the medical setting. This represents the heuristic advantage of CART analysis. This paper is not intended to furnish new information, but to improve already existing knowledge, using a different instrument.

How is the classification tree used to predict 'post-consensus clinicians' diagnosis of asthma'? From the root to the base of the answer to each predictor variable, we move along the branches of the tree until we arrive at a terminal node – or 'leaf' – of the tree which allows us to define the subject as 'asthmatic' or 'non-asthmatic', according to whether the 'leaf' is shaded or not shaded. Perhaps an example will help to understand the rule better. We wish to know if a subject should be allocated as asthmatic or non-asthmatic class using the decision tree proposed in Figure 1. First, we observe the subject's answer to the question 'ever asthma'. Supposing that the answer is 'Yes' we move from the root to the left branch. Since this is a shaded leaf we conclude that the subject will be classified as asthmatic. On the other hand, if the answer is 'No' we would cross to the right branch, which introduces an intermediate node (a circle) followed by a bifurcation. In this case it is necessary to continue along the branches of the tree and answer a new question on 'current asthma'. Supposing that the subject's answer is 'No' we would continue along the right branch where an unshaded leaf is met and we conclude that the subject would belong to the 'non-asthmatic' class. Conversely, if the subject answers 'Yes' to 'current asthma', we would continue on to the left branch and meet a second intermediate node, or rather, a new bifurcation. This involves answering a new question about 'shortness of breath'. If the subject responds 'Yes' we would meet a shaded leaf and conclude he is asthmatic, while if he answers 'No' we would meet the end of the right branch where there is another circle. So it would be necessary to continue along the tree.

The decision tree underscores the fact that among the items of the questionnaire the most important one results 'ever asthma' or self-awareness of the disease. This represents the first split in the diagram tree; furthermore, a positive response to this question involves immediate allocation of a subject to the 'asthmatic' class, independently of the other variables. This subgroup represents more than 10% of the sample. The questions about asthma here were more specific, while, in Burney et al.' paper, they were less sensitive of wheeze and breath ones [24]. The variable 'current asthma' finds that almost 82% of the remaining 90% 'non-asthmatic' subjects are truly 'non-asthmatic'. In other words, we can say that subjects answering negatively to 'ever asthma' and to 'current asthma' are 'non-asthmatics', independently of the subsequent predictors in the classification tree. Moreover, of the remaining 18% of subjects positive for 'current asthma', more than 65% of them are 'asthmatic' if they are positive for at least one more of the subsequent, hierarchically lower, predictors.

Attention must be drawn to the fact that the above mentioned 'proportions' are not 'prevalences' of the disease in the population, rather they are a measure of predictor importance towards the other predictors in the decision rule. The high percentage of cases indicated by the clinicians' diagnosis must be considered in relation to the nature of the sample selected of stage II. Not only was this not a random sample, but it also contained a greater number of symptomatic subjects (resulting from stage I of survey) than non-symptomatic ones. Furthermore, as shown in stage I, responders were more symptomatic than non-responders and this was even more true in stage II. Therefore, it is obvious that the prevalence of a 'post-consensus clinicians' diagnosis of asthma' in a 'filtered' sample of population subjects would differ from the true proportion of disease; in any case the objective of this study was not to estimate the prevalence of asthma.

We propose cutting the last two leaves in the classification tree, that is eliminating the predictor 'breathless after effort', which leads to a more interpretable tree. The new tree evolves as a hybrid of the CART pruning process and clinical expertise. We noted an inhibiting effect between 'current asthma' and 'breathless after effort' (one of the components of 'current asthma') in the last terminal nodes. CART classifies a portion of subjects positive for 'current asthma' as non-asthmatic if they are positive for 'breathless after effort'. In fact, of the 59 filtered subjects who responded positively to being 'breathless after effort', only two were asthmatic. By stopping the classification tree of the preceding bifurcation, it follows that the negative branch 'wheezing and breathlessness' finishes in an unshaded leaf (risk ratio = 1), so, there will be six leaves, four shaded and two unshaded in the new decision tree. The goodness-of-fit of new tree is supported by the resubstitution measures (sensitivity = 84.6%, specificity = 93.3%, accuracy = 92%).

Our CART analysis confirms the validity of asthma questionnaire items for measuring 'asthma' in epidemiological surveys as underlined by Jenkins et al. [12]. When the sensibility produced by the CART analysis (rule B or the new rule as mentioned above) is compared with that of the single predictors, a considerable improvement is noted. In fact, the percentage of true sick (88% or 85%) is very good, if we remember that for each individual predictor it is less then 62%. A review of the literature, with respect to Italian studies, confirms that the sensitivity and specificity of the classification tree are both good with respect to those (79.2% and 88.6% for the best index, respectively) found by Bruschi et al. [25] in the general population of a small Lombardy town for different indices of bronchial hyperactivity, when the clinical diagnosis was utilised as the gold-standard. Furthermore, De Marco et al. [19] compare single questions, combination of questions, and combination of questions and tests using same ECRHS data and post-consensus diagnosis as gold standard. Only the combination named past-year symptoms achieved good results (sensitivity = 82.9% and specificity = 86.7%) using an optimistic evaluation of classifier procedure (the 'resubstitution method'), vice versa we utilised the 10-fold cross-validation samples method.

The importance of symptoms and bronchial challenge testing in respiratory research is indicated in this paper, along with the fact that BHR cannot reliably or precisely separate asthmatic from non-asthmatic subjects in the general population. In fact, our classification tree takes into account the possibility that a subject who does not exhibit BHR can be asthmatic, while having only self-awareness of asthma or at least asthma-like symptoms.

The main findings of the present paper permit us to draw some important conclusions. First, as we have already mentioned, these results are based on our belief that the selection of a 'post-consensus clinicians' diagnosis of asthma' as the gold standard is well grounded, similarly to Jenkins et al. [12] and Cerveri et al. [9]. So it has been possible to evaluate an epidemiological instrument for assessing the presence of asthma by examining simple clinical tests (pulmonary function and allergy tests), past-year asthma symptoms and self-awareness of asthma (the latter probably depending on a previous diagnosis made by a general practitioner).

Second, of all the information that was collected, the symptomatological questionnaire was the most useful for identifying asthma cases in epidemiological studies, and in particular the question 'Have you ever had asthma?'. The symptoms during the past year concerning the working definition of 'current asthma' are also important: 'Being breathless when a wheezing noise was present', 'Wheezing or whistling in the chest in the absence of cold', and 'Having an unexpected attack of breathlessness after intense effort'. Another helpful tool is the 'BHR' challenge procedure, necessary in the definition of already mentioned 'current asthma' and, hierarchically lower, the atopy (as measured by a positive skin test to any one of a panel of allergens).

These findings are of special interest if one considers the cost of epidemiological studies that involve medical examinations and numerous instrumental tests. Naturally this holds true for identifying a disease state and not for determining which variable is predictive of developing the disease. The latter objective can greatly benefit from the results of ECRHS follow-up study conducting to years later on this same group of patients.

Acknowledgements

The authors wish to thank the 1991–1993 Italian group of ECRHS for the research data. The members of ECRHS-Italy group are the following: *University of Verona*: de Marco R, Lo Cascio V, Campello C, Rossi F, Biasin C, Cannistrà A, Cenci B, Destefani E, Ferrari M, Girotti M, Lamprotti G, Martini C, Olivieri M, Poli A, Tardivo S, Verlato G, Villani A, Zanolin ME; *University of Pavia*: Marinoni A, Cerveri I, Alesina R, Basso O, Berrayah L, Brusotti R, Fanfulla F, Moi P, Zoia MC; *University of Turin*: Bucca C, Romano C, Aime M, Cerutti A, Chiampo F, Gallo G, Rola, Sulotto F; *ASL 42 Pavia*: Casali L, Fratti C, Karytinos P; *ASL 7 Torino*: Bugiani M, Arossa W, Caria E, Carosso A, Castiglioni G, Piccioni P.

Appendix: Assessment and options of decision rules of CART, LDF and MLP

The decision rules of CART, LDF and MLP methods are concerned with the relationship between the group membership label of the post-consensus diagnosis (the response variable) and the set of predictor variables (asthma-like symptoms, respiratory function testes, allergy test, etc.). The discrimination procedures chosen do not assume any particular parametric form of the multivariate distribution of the two populations (non-case vs. case).

The measure of the quality of the decision rules is determined through some misclassification errors. Assuming a distributional probability model of the data that determines theoretical values of the misclassification errors can draw optimal decision rules, which minimise these errors. If no distributional assumption is required, as in our case, an indication of the misclassification errors can be obtained directly from the data. It is always possible to build the following 2×2 table (called a confusion table):

If group = 1 and group = 2 designate the groups of the healthy (non-case) and the sick (case) subjects

		Predicted class 1	2	
True class	1	n_{11}	n_{12}	$n_{1.}$
	2	n_{21}	n_{22}	$n_{2.}$
		$n_{.1}$	$n_{.2}$	n

respectively, the usual quantities of evaluation of a diagnostic test can be determined as follow:

$$\text{Specificity} = n_{11}/n_{1.}; \quad \text{Sensitivity} = n_{22}/n_{2.};$$
$$\text{Accuracy} = (n_{11} + n_{22})/(n_{1.} + n_{2.})$$

from which we get the type I (false positive), $\alpha = n_{12}/n_{1.} = (1 - \text{specificity})$ and type II (false negative), $\beta = n_{21}/n_{2.} = (1 - \text{sensitivity})$ misclassification errors and the average (total) misclassification error, $(n_{1.}\alpha + n_{2.}\beta)/(n_{1.} + n_{2.}) = (1 - \text{accuracy})$.

These misclassification errors are called 'apparent errors' and this empirical method is called the 're-substitution method' (because the individuals in the sample are used both to draw the rule of assignment and to measure the performance of discrimination). In general, the 'apparent error' overestimates the true misclassification error; therefore different resamplings and cross-validation methods have been proposed for obviating this problem [22, 23], for examples:

- *split-samples*: the data set is divided into two parts. On the first data set, called the 'training sample', the allocation rule is obtained and on the second, called the 'validation sample', the misclassification errors or 'actual errors' are determined;
- *K-fold cross-validation samples*: the sample is divided into K equal-sized pieces; the allocation rule is calculated on a partial sample formed by $K-1$ parts, and the misclassification errors on the remaining part. This is repeated for $k = 1, 2, \ldots, K$. The K misclassification errors are then averaged to give cross-validation measures of actual errors;
- *bootstrap samples*: a series of B random samples of n elements is generated by sampling with replacement from the data set. The average of B 'apparent errors' that results when the data set is the bootstrap sample is the simple bootstrap estimate of actual errors. A more refined estimate is also possible [26].

The present study utilised the 10-fold cross-validation samples method both to develop decision rules and to compare the different classification techniques.

For each kind of classifier technique, three classification rules (A, B and C) were applied, using different options concerning the 'misclassification risk ratio' $= \pi_2 c_2/\pi_1 c_1$, where π_1 and π_2 are the prior probabilities that the subjects come from class 1 = non-asthmatic or class 2 = asthmatic, respectively; and c_1 and c_2 are the penalties or cost functions due to misclassification of a 'non-asthmatic' subject as 'asthmatic', or vice versa: misclassifying a class 1 subject incurs a 'cost' of c_1 and misclassifying a class 2 subject incurs a 'cost' of c_2, respectively.

For CART rule A, the prior probabilities are equal to the observed prevalences of post-consensus asthma in the sample: $\pi_1 = p_1$ and $\pi_2 = p_2$; while the penalty for misclassifying an asthmatic subject as non-asthmatic is twice of that for misclassifying an non-asthmatic subject as asthmatic: $c_1 = 1$ and $c_2 = 2$. In rule B we assign: $\pi_1 = p_1$ and $\pi_2 = p_2$, but 10 penalty points are charged for misclassifying subjects in the true 'asthmatic' class and one penalty point is charged for misclassifying subjects in the true 'non-asthmatic' class: $c_1 = 1$ and $c_2 = 10$. In rule C we imposed equal prior probability and cost functions: $\pi_1 = \pi_2 = ½$ and $c_1 = c_2 = 1$.

In LDF the cost functions are related by $c_1 + c_2 = 1$. So we assign in rule A: $\pi_1 = p_1$, $\pi_2 = p_2$ and $c_1 = 0.09$, $c_2 = 0.91$, i.e. $c_2/c_1 = 10$; rule B: $\pi_1 = \pi_2 = ½$ and $c_1 = 0.09$, $c_2 = 0.91$ and in rule C we imposed equal prior probability and cost functions: $\pi_1 = \pi_2 = ½$ and $c_1 = c_2 = 1$. While for MLP we use in rule A–B: $\pi_1 = \pi_2 = ½$ and $c_1 = c_2 = 1$; in rule C we assign a prior probability equal to an hypothetical prevalence of asthma in the population: $\pi_1 = 0.95$ and $\pi_2 = 0.05$ and we impose equal cost function: $c_1 = c_2 = 1$.

In addition, we defined two different criteria for assessing the performance of two-class decision rules [27]:

- *minmax error* rate *criteria*: for screening purposes it is better to have a classification rule such that the maximum rate of each misclassification type must be minimised; specifically, we will choose a two-class rule so that the sensitivity is equal (or almost equal) to the specificity. It immediately follows that the false negative rate is equal (or almost equal) to the false positive rate. This measure is the complement of Youden's index [28];
- *minmax error* number *criteria*: for screening purposes it is better to give a classification rule such that the numbers being put into each class are equal (or almost equal) to the numbers known to have come from that class, that is the two marginal totals of the confusion table are equal (or almost equal). It immediately follows that the false negative number is equal (or almost equal) to the false positive number.

If we have equivalence with respect to previous criteria we chose the rule with the greatest sensitivity. If sensitivity tends to one then negative predictive value tends to one (for any prevalence value). This means that the subjects identified as 'non-asthmatic' by the decision rule are certainly the healthy subjects, while those identified as 'asthmatic' are all the sick ones plus a quota of healthy subjects (false positives). Therefore, if a decision rule is used for a two-phase diagnosis only the group of positive subjects (generally much smaller than that of the negatives) will be

submitted, in a second occasion, to a more careful survey to eliminate the false positive subjects.

References

1. Sears MR. Descriptive epidemiology of asthma. Lancet 1997; 350(2s): 1–4.
2. Ciba Foundation Guest Symposium. Terminology definition and classification of chronic pulmonary emphysema and related conditions. Thorax 1959; 14: 286–299.
3. American Thoracic Society. Definitions and classifications of chronic bronchitis, asthma and pulmonary emphysema. Am Rev Respir Dis 1962; 85: 762–768.
4. World Health Organisation. Epidemiology of chronic non-specific respiratory diseases. Bull WHO 1975; 52: 251–259.
5. Gross NJ. What is this thing called love?–Or, defining asthma (editorial). Am Rew Respir Dis 1980; 121: 203–204.
6. Toelle BG, Peat JK, Salome CM, Mellis CM, Woolcock AJ. Toward a definition of asthma for epidemiology. Am Rev Respir Dis 1992; 146: 633–637.
7. Global strategy for asthma management and prevention – NHLBI/WHO Workshop report (March 1993). Bethesda: National Institute of Health, National Heart Lung and Blood Institute. Publication No 95-3659, 1995.
8. Burr ML. Diagnosing asthma by questionnaire in epidemiological surveys (editorial). Allergy 1992; 22: 509–510.
9. Cerveri I, de Marco R, Bugiani M, et al. Can clinical judgement be a gold standard for identifying bronchial asthma in epidemiological studies? Interobserver agreement and contribution of specific diagnostic procedures. Monoldi Arch Chest Dis 1998 (submitted).
10. Lilienfeld AM, Kordan B. A study in the interpretation of chest X rays in the diagnosis of lung-cancer. Cancer Res 1966; 26: 2145–2147.
11. Feinsten AR, Geilfman NA, Yesner R, Auerbach O, Hackel DB, Pratt CB. Observer variability in histopathologic diagnosis of lung cancer. Am Rev Respir Dis 1970; 101: 671–684.
12. Jenkis MA, Clarke JR, Carlin JB, et al. Validation of questionnaire and bronchial hyperresponsiveness against respiratory physician assessment in diagnosis of asthma. Int J Epidemiol 1996, 25(3): 609–616.
13. Burney PGJ, Luczynska C, Chinn S, Jarvis D, for the European Community Respiratory Health Survey. The European Community Respiratory Health Survey. Eur Respir J 1994; 7: 954–960.
14. The European Community Respiratory Health Survey. 'Medicine and Health', European Commission, Directorate-General XIII, Office for Official Publications, L-2920 Luxembourg, 1994.
15. De Marco R, Verlato G, Zanolin E, Bugiani M, Drane JW. Nonresponse bias in EC Respiratory Health Survey in Italy. Eur Respir J 1994; 7: 2139–2145.
16. European Community Respiratory Health Survey – Italy. Prevalence of asthma and asthma symptoms in a general population sample from northern Italy. Allergy 1995; 50: 755–759.
17. Verlato G, Cerveri I, Villani A, et al. Evaluation of methacholine dose-response curves by linear and exponential mathematical models: Goodness-of-fit and validity of extrapolation. Eur Respir J 1996; 9: 506–511.
18. Cerveri I, Zoia MC, Bugiani M, et al. Inadequate antiasthma drug use in the north of Italy. Eur Respir J 1997; 10: 2761–2765.
19. De Marco R, Cerveri I, Bugiani M, Ferrari M, Verlato G. An undetected burden of asthma in Italy: The relationship between clinical and epidemiological diagnosis of asthma. Eur Respir J 1998; 11: 599–605.
20. Pepys J. Atopy: A study in definition (editorial). Allergy 1994; 49: 397–399.
21. American Thoracic Society. Evaluation of impairment/disability secondary to respiratory disorders. Am Rev Respir Dis 1986; 133(6): 1205–1209.
22. Hand DJ. Construction and Assessment of Classification Rules. New York (US): Wiley, 1997.
23. Morineau A. Introduction à SPAD, Version 3.01. Saint-Mandé (France): CISIA, 1996.
24. Burney PGJ, Chinn S, Britton JR, Tatterfield AE, Papacosta AO. What symptoms predict the bronchial response to histamine? Evaluation in community survey of the bronchial symptoms questionnaire (1984) of the International Union Against Tuberculosis and Lung Disease. Int J Epidemiol 1989; 18(1): 165–173.
25. Bruschi C, Cerveri I, Zoia MC, Maccarini L, Grassi M, Rampulla C. Bronchial responsiveness to inhaled methacoline in epidemiological studies: Comparison of different indices. Eur Respir J 1989; 2: 630–636.
26. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman & Hall, 1993.
27. Hand DJ. Screening vs prevalence estimation. Appl Statist 1987; 36: 1–7.
28. Armitage P, Berry G. Statistical Methods in Medical Research. Oxford: Blackwell Scientific, 1994.

Address for correspondence: Mario Grassi, Dipartimento di Scienze Sanitarie Applicate e Psicocomportamentali, Università degli Studi di Pavia, Via Bassi, 21, I27100, PAVIA, Italy
Phone: +39-382507535; Fax: +39-382507570
E-mail: mariog@unipv.it