ELSEVIER

Using a Neural Network to Screen a Population for Asthma

SYBIL HIRSCH, PHD, JONATHAN L. SHAPIRO, PHD, MICHAEL A. TUREGA, PHD, TIMOTHY L. FRANK, MD, MRCGP, ROBERT McL. NIVEN MD, FRCP, AND PETER I. FRANK, MD, FRCGP

PURPOSE: To use a neural network to rank a population according to individual likelihood of asthma based on their responses to a respiratory questionnaire.
METHODS: A final diagnosis of asthma can be made only after full clinical assessment but limited resources make it impossible to offer this to complete populations as part of a screening programme. Prioritisation is required so that review can be offered most promptly to those most in need. A stratified random sample of 180 from 6825 respondents to a community survey underwent clinical review. They were categorised according to likelihood of asthma by three independent experts whose opinions were combined into a single probability label for each patient. A neural network was trained to relate questionnaire responses to probability labels. The trained network was applied to the whole community to produce a ranking order based on likelihood of asthma. A screening threshold could then be set to correspond to available resources, and patients above this level with no recorded evidence of asthma diagnosis could be assessed clinically. Using the known probability labels from the training set, it was possible to derive the expected proportion of true asthmatics in any set of patients.
RESULTS: If the screening threshold had been set to capture the top 10% of the ranked population (n = 683), then 239 patients above this threshold had no evidence of diagnosis and would need assessment. Of these, it would be expected that 74% would have the diagnosis confirmed.
CONCLUSIONS: This approach allows prioritisation of a population where resources for diagnostic examination are limited.
Ann Epidemiol 2001;11:369–376. © 2001 Elsevier Science Inc. All rights reserved.

KEY WORDS: Asthma, Bayes Theorem, Neural Network (Computer), Probability, Mass Screening

INTRODUCTION

Recent studies have indicated that the prevalence of asthma is increasing (1–3) and it may be under-diagnosed and under-treated (4, 5). It is therefore likely that there are individuals in the community who would benefit from treatment but do not receive it because their condition is not known to the medical services. This possibility gives rise to particular concern in view of the theory that airway remodelling and possible fixed asthma can ensue from failure to treat asthma airway inflammation (6).

Final diagnosis of asthma can be made only after a full clinical assessment. However, it would not be practical to offer this to entire populations in an attempt to identify those individuals with unrecognised asthma. A screening technique to identify a smaller target group of possible asthmatics has been attempted as part of the Wythenshawe Community Asthma Project (WYCAP) (7, 8), a long term prospective study examining the natural history of asthma in a community population in South Manchester, UK. The technique used a simple patient-completed postal questionnaire (Figure 1) and scored the number of positive responses to questions believed to be indicative of asthma. Initially an arbitrary screening threshold of four positive responses out of six key questions was used to identify individuals who were more likely to have asthma. However, concern was raised that at this threshold the number screening positive may prove to be so great that prompt clinical evaluation could not be offered to all. It was therefore important to develop a more refined method of prioritisation to identify those with the greatest likelihood of benefiting from intervention.

This paper describes a new approach in which the questionnaire responses for each patient were processed using a

From the General Practice Research Unit, North West Lung Research Centre, Wythenshawe Hospital, Manchester, United Kingdom (S.H., T.L.F., R.M.N., P.I.F.); Department of Computer Science, University of Manchester, Manchester, United Kingdom (S.H., J.L.S.); and Computation Department, UMIST, Manchester, United Kingdom (M.A.T.).
Address reprint requests to: Dr. Sybil Hirsch, General Practice Research Unit, North West Lung Research Centre, Wythenshawe Hospital, Southmoor Road, Manchester M23 9LT, United Kingdom.
Received February 17, 2000; revised October 9, 2000; accepted October 14, 2000.

© 2001 Elsevier Science Inc. All rights reserved.
655 Avenue of the Americas, New York, NY 10010
1047-2797/01/$–see front matter
PII S1047-2797(01)00233-2

Selected Abbreviations and Acronyms
WYCAP = Wythenshawe Community Asthma Project
MLP = Multi-layer perceptron

more sophisticated statistical technique known as an artificial neural network (9–12). This technique uses complex relationships between questions in order to predict each respondent's probability of asthma. The population could then be ranked according to these probability predictions and a screening threshold set to reflect the percentage of patients the medical services can afford to review. Patients above this cut-off with no evidence of asthma in their medical records could be given high priority for early clinical assessment.

The Neural Network

The neural network is a computerised software programme, which can be used to develop a non-linear statistical model in which there are many parameters, called weights. These weights control the output of the model (probability of asthma) given the input (questionnaire responses). The type of network used was a fully connected feed forward back propagation multi-layer perceptron (MLP) (13–15). The network can be visualised as interconnected nodes arranged in layers, as shown in Figure 2. The nodes in the first layer represented the questionnaire responses and there was a single output representing the probability of asthma. In addition there was a 'hidden' layer, whose nodes allowed the network to develop the more complex, non-linear models commonly required to describe biological data.

The complexity of the model is related to the number of hidden nodes. One aspect of the training process was therefore to try networks with different numbers of hidden nodes in order to determine an appropriate model complexity. In each case, the network was 'trained' by repeatedly presenting it with example data of questionnaire responses from individuals of known asthmatic status. The connections between nodes represented the weights of the model and during training the network's iterative algorithm gradually modified the weights. Training was stopped when the network model had the optimal fit. The final trained network system was then used to predict on an individual basis the probability of asthma in individuals in a wider population.

METHODS

Training Set

The training set was created from the 1995 WYCAP survey. In this survey, a postal screening questionnaire based on the European Respiratory Health Survey Questionnaire (16) was sent to all adults (16 years and over) in two neighbouring South Manchester general practice populations. In order to allow more accurate determination of the response patterns of both asthmatic and non-asthmatic responders it was necessary to have an asthma-enriched sample. Using clinical intuition, the respondents were therefore divided into five strata based on the number of positive responses [0, 1–3, 4, 5, and 6] to 6 key questions. These were questions which were believed to be important in the detection

Please tick the appropriate box.

1.	What is your date of birth?		
2.	Are you	FEMALE ☐	MALE ☐
3.	Have you had wheezing or whistling in your chest at any time **in the last 12 months?** **IF 'NO'** GO TO QUESTION 4, **IF 'YES'**:	NO ☐	YES ☐
	3.1 Have you been at all breathless when the wheezing noise was present?	NO ☐	YES ☐
	3.2 Have you had this wheezing or whistling when you did not have a cold?	NO ☐	YES ☐
4.	Have you woken up with a feeling of tightness in your chest **in the last 12 months?**	NO ☐	YES ☐
5.	Have you been woken by an attack of shortness of breath at any time **in the last 12 months?**	NO ☐	YES ☐
6.	Have you been woken by an attack of coughing at any time **in the last 12 months?**	NO ☐	YES ☐
7.	Have you had an attack of asthma in the last 12 months?	NO ☐	YES ☐
8.	Are you currently taking any medicine for asthma? (including inhalers, aerosols or tablets)	NO ☐	YES ☐
9.	Has any person in your family (parents, grandparents, sisters, brothers, or your children) had asthma?	NO ☐	YES ☐
10.	Have you ever had hayfever or eczema?	NO ☐	YES ☐
11.	How many cigarettes do you smoke each day? .....per day		
12.	How many other adults live in your house?	☐	
	How many of these adults smoke?	☐	

THANK YOU FOR YOUR HELP
PLEASE RETURN THIS FORM TO US IN THE REPLY-PAID ENVELOPE

FIGURE 1. The Screening Questionnaire. See Figure 1 on Page 27 of original manuscript.

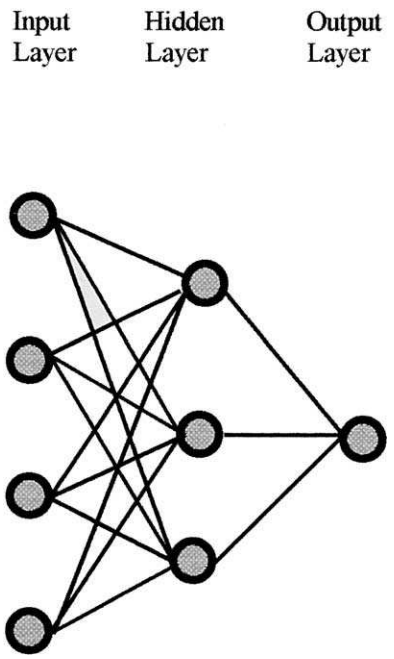

FIGURE 2. Diagrammatic representation of the Multi-Layer Perceptron Neural Network. The circle represents a node and the line represents a connection between nodes. For this study, there were 16 input nodes representing questionnaire responses and a single output node representing probability of asthma.

of an asthmatic, namely questions relating to symptoms in the last 12 months (wheezing, chest tightness, shortness of breath, night cough), family history of asthma and associated conditions of hay fever or eczema (Figure 1). A random sample of patients was selected from within each stratum and the target numbers were weighted in favour of the higher scoring strata aiming to provide approximately 50% asthmatics overall.

The selected individuals were contacted and invited for a clinical assessment at the North West Lung Research Centre. Assessment comprised a full history, physical examination, spirometry, reversibility of symptoms with β-2 agonists, electronic peak flow diaries, and skin prick testing to five common allergens. This information was sent separately to three respiratory physicians. With no knowledge of the postal questionnaire responses, each expert allocated each of the reviewed individuals to one of three diagnostic categories, greater than 90%, 50–90%, and less than 50% probability of asthma respectively, giving 27 possible permutations of expert opinions. The use of probability interval labels was an attempt to standardise the more subjective interpretation of textual labels such as probable, possible and unlikely asthma.

The postal questionnaire responses from these individuals were used as the basis of the training set. For each person there was a set of attributes from the questionnaire responses for age, gender, respiratory symptoms, recent asthma attack, asthma medication, family history of asthma, associated conditions and smoking habits. However, there was no label of asthmatic status. Instead, there was the set of three expert diagnoses into one of the three classes.

The development of the prioritisation system entailed three steps. First, the expert diagnoses for the reviewed individuals were used to label their questionnaires with an estimate of the probability of asthma of the respondent; second these labelled questionnaires were used to train the network; third the trained network was applied to the wider population in order to assess its effectiveness as a tool for finding hitherto undiagnosed asthmatics.

Probability Labels

In order to provide the required training labels it was first necessary to combine the three often differing expert opinions into a single probability label for each individual in the training set (17). In other words, a model was required which allowed the calculation of the probability of asthma for the individuals, given their three expert diagnoses. First, the theoretical foundations of this model will be described, followed by the practical details of how it was applied to label the training examples.

Theoretical Foundations of the Labelling Model

In mathematical terms, each training label is the value of a function P defined by $P(a \mid R, S, T)$, where a denotes the condition of asthma, and R, S, and T are variables which take the values 1, 2, or 3 denoting the diagnoses of the three experts respectively into one or other of the probability classes 1 (> 90%), 2 (50–90%), or 3 (<50%).

In order to determine the label value corresponding to each of the 27 permutations, standard probability theory was applied to express the label in terms of some basic parameters. Two of these were the probability of asthma and non-asthma in the sample, denoted p_0 and q_0, respectively. There was one series of parameters describing how the experts classified asthmatics, for example, the probability that expert A classifies an asthmatic as class 1 and the probability that expert C classifies an asthmatic as class 2, denoted p_{A1} and p_{C2}, respectively.

Finally, there was a series describing how experts classify non-asthmatics, denoted q_{A1}, q_{A2}, q_{A3}, and so on. Altogether, for three experts (A, B, and C) there were a total of 20 parameters. However, only 13 of these were independent because each person is either asthmatic or non-asthmatic and each expert must make some diagnosis (Appendix 1).

If the values of the 13 independent parameters were known, it would be possible to calculate the probability of asthma for each of the reviewed individuals from their three diagnoses. For example, applying Bayes Rule (18) and rules of probability, if the experts A, B, and C classified a person in classes 2, 3, and 1, respectively, the label would be

$$P(a|2,3,1) = \frac{p_{A2}p_{B3}p_{C1}p_0}{p_{A2}p_{B3}p_{C1}p_0 + q_{A2}q_{B3}q_{C1}(1-p_0)} \quad (1)$$

and so on, for all individuals in the subsample.

Practical Details of the Labelling Method

Unfortunately, the values of these model parameters were not known and could not be directly measured due to the fundamental problem that there were no known asthmatics or known non-asthmatics in the training set. A Bayesian approach was therefore adopted which allowed estimation of label values by combining two sources of information. The first was an intuitive belief in the kinds of values the parameters might take, referred to in Bayesian terms as the prior probability distribution, denoted *P(parameters)*. Although its details were unknown, it was possible to specify the prior in a general way based on two conditions, which defined a common-sense belief in some minimum level of competency of experts to diagnose asthma.

First, it was assumed that an expert was better than a guesser, making fewer mistakes than just guessing, say, the most likely class. Second, it was assumed that even though an expert classification could not be taken at face value it did give some broad general information. For example, the expert is more likely to classify an asthmatic than a non-asthmatic into classes 1 or 2 ($>$ 50% probability of asthma). These conditions were expressed mathematically as a series of constraints (Appendix 2).

The second source of information required for the Bayesian methods was the evidence provided by the 180 expert classification patterns {111,112,121, . . . 333}, in particular, the numbers $\{N_{111}, N_{112}, N_{121}, \ldots, N_{333}\}$ of each type. This evidence was expressed mathematically in the form of a so-called likelihood function, which represented the probability of those 180 patterns occurring for any particular set of parameter values. The likelihood function was denoted *P(data | parameters)* (Appendix 2).

The purpose of specifying the prior and likelihood functions mathematically was to combine them according to Bayes Rule into a specification of the posterior probability function, denoted *P(parameters | data)*. This is represented in the equation

$$P(parameters \mid data) = \frac{P(data \mid parameters)\, P(parameters)}{P(data)} \quad (2)$$

where *P(data)* was a normalisation factor to ensure that the parameter values, which were probabilities, were in the required range 0–1. The posterior function, *P(parameter | data)*, represented our final belief after taking into account the evidence of the expert classification data in the kinds of values the parameters might take and their likelihood of occurrence. If this posterior function could be quantified, in other words if these values and their probability of occurrence could be determined, the relationship in Equation 1 could be applied to specify a corresponding distribution of label values This distribution would be 27-dimensional representing the probability of asthma corresponding to the 27 expert classification patterns. It would then remain to choose from this distribution which particular set to use as training labels. In this situation a maximum likelihood approach is often adopted whereby the most commonly occurring value for each of the 27 components would be chosen. However, there was no reason to believe that the label distribution here would show clear 'winners' and so the mean value for each of the 27 components was considered more appropriate.

Direct calculation of these mean values was not possible since this would require the evaluation of multi-dimensional integrals, which cannot be computed analytically. Instead the mean label values were estimated using Monte-Carlo sampling methods. Using an algorithm known as the Metropolis algorithm (19, 20), it was possible to use the prior and likelihood functions to select a sample of points from the posterior parameter distribution *P(parameters | data)*. The sample produced, known as a Markov chain (19, 20, 21), consisted of 60000 points each of which was a set of 13 possible parameter values. It was then a simple matter to use the relationship in Equation 1 to derive a corresponding sample from the label distribution and use the mean value of the 27 components to label the training examples.

Although there was no external standard against which to judge the accuracy of the label values, they were assumed to be the best estimates of probability of asthma and consequent ranking of the reviewed individuals. These labels were therefore used as the baseline 'gold-standard' against which to train the network and assess its effectiveness in ranking the population.

Training the Network

Next, the network was trained to relate the questionnaire responses from these reviewed individuals to their probability labels . Most of the responses were presented to the network as '0' for 'No' and '1' for 'Yes'. There were two continuous variables, date of birth and smoking, which were converted from continuous to categorical values resulting in four age bands [16–34, 35–54, 55–74, and $>$ 75 years, respectively] and the two categories 'Smoking' and 'Non-smoking'. In all, the eleven questions resulted in 16 network inputs.

In order to make best use of the small number of labelled examples, training was performed using cross-validation (22) in which the set of 180 labelled questionnaires was partitioned into 18 randomly selected test sets each with 10 examples. For each test set, the training process was carried out on the remaining 170. Each set of 170 examples was randomly divided into two sets known as training and optimisation sets, containing 120 and 50 examples, respectively.

The 120 training examples were repeatedly presented to the network and after each pass the training algorithm (back propagation) modified the weights in a direction, which reduced the difference between network outputs and training labels. The purpose of the 50 optimisation examples was to provide a criterion for stopping training before the model became over-fitted to the 120 training examples. The criterion used was that the mean square error between network outputs and optimisation labels was a minimum.

This routine was repeated for networks of various complexities (1, 2, 3, and 4 nodes in the single hidden layer). For each of the 18 partitions, the network which produced the minimum mean square error on the 50 optimisation examples was deemed the 'best' from that partition. This was then used to predict the asthma scores of the remaining 10 test examples to provide an independent assessment of how the network would perform in the wider population. This process resulted in 18 different networks (Figure 3). The accuracy of the network in ranking each of the 18 test sets was measured by the Spearman rank order correlation coefficient (23). This method of cross-validation allowed all 180 examples to be used in testing at a time when they had not been involved in training or optimisation. In addition, the 18 network models could be applied as a 'committee' to the wider community, avoiding the problem of using one, possibly unrepresentative network model. The mean Spearman value over all 18-test sets was used as an estimate of how accurately the network would rank the wider community (24).

The Wider Population

The third step was to run the 18 'committee' networks on the screening questionnaires from the whole of the study population ($n = 6825$). For each respondent, the estimated probability of asthma was taken as the mean value of all 18 network outputs. The respondents were listed according to these mean value probability estimates thus providing the desired population ranking.

In order to use the system for a screening programme, individuals above a chosen threshold with no corroborative evidence of asthma in their general practice medical records would be selected for clinical assessment. For example, in the UK this group could easily be identified from the complete primary and secondary care data held by their general practitioner. The threshold would depend on the resources available. The effectiveness of the ranking method was assessed by estimating the number of individuals in this selected group, for who a diagnosis of asthma would be expected to be confirmed. In order to avoid over-estimation of these expected numbers it was necessary to adjust for the known differences in prevalence between the strata in the training set and those in the wider population. This was achieved by using the known probability labels in the training set to estimate the expected probability of asthma of each individual in the population. The sum of these expected probabilities was used as an estimate of the expected prevalence of asthma diagnosis in the population sub-group selected for clinical assessment (17, 25).

RESULTS

In the 1995 survey, 10,429 patients were sent a postal screening questionnaire, of whom 7583 responded (72.7%). The ages ranged from 16 to 96 years and 45% were from males. Approximately one third of the respondents were smokers. A total of 3146 individuals (41%) reported a family history of asthma, and 1839 individuals (24%) said they had an associated condition such as eczema or hay fever. Thirty percent reported wheezing in the last twelve months. A stratified random sample of four hundred and twenty patients was selected for clinical review. The general practitioner succeeded in contacting 240, of whom 202 finally attended for the examination. For the present analysis, incomplete questionnaires were excluded, leaving 6825 screening questionnaires, of which 180 were from reviewed individuals.

There was considerable disagreement between the experts, kappa = 0.23 (26), as to the correct class for some of the patients, with all three experts agreeing the same probability interval for only 32% of the labelled set. In 10% of the labelled set, each expert chose a different probability interval.

Combination of the 27 permutations of expert opinions resulted in probability labels ranging from 0.0013 to 0.9991 (Table 1). The 180 labelled examples were used to train the

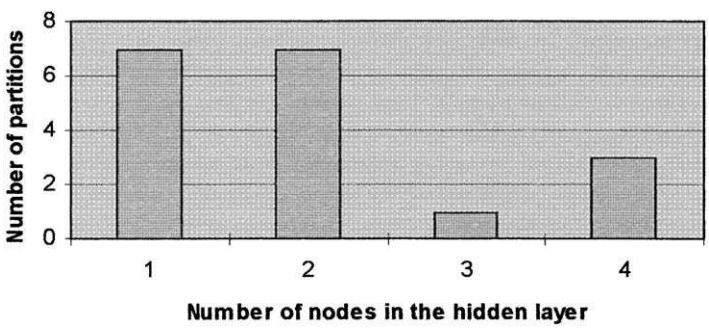

FIGURE 3. Distribution of multi-layer perceptron architectures from the optimised cross validation partitions. Most partitions optimised to networks with 1 or 2 nodes in the hidden layer.

network by relating them to the postal questionnaire responses. The mean Spearman rank order correlation coefficient over all 18 test sets, estimating network accuracy in ranking the wider community, was 0.66.

The trained network committee was used to rank all 6825 study questionnaires according to their individual probability estimates. The expected numbers of asthmatics corresponding to various screening thresholds are shown in Table 2. For example, if this threshold were set to include the top 10% of the population as ranked by the neural network, 683 individuals would lie above the threshold. Of these, 239 had neither diagnosis of asthma nor recent prescription for asthma medication in their general practice records and would therefore be selected for review. Seventy four percent of these reviewed individuals would be expected to be diagnosed asthmatic on clinical examination. If the threshold were set to include the top 5%, the expected number of false positives would decrease and 88% of the 68 individuals invited would be expected to have the diagnosis confirmed.

TABLE 1. Frequencies of reviewed individuals and probability labels with standard deviations for all 27 permutations of expert opinions (n = 180)

Expert Diagnoses[a]			Number of patients	Percentage of patients	Probability label value *(standard deviation)*
A	B	C			
1	1	1	15	8.3	0.9991 *(0.0009)*
1	1	2	2	1.1	0.9948 *(0.0046)*
1	1	3	1	0.6	0.9644 *(0.0464)*
1	2	1	13	7.2	0.9880 *(0.0090)*
1	2	2	4	2.2	0.9386 *(0.0382)*
1	2	3	4	2.2	0.7181 *(0.1478)*
1	3	1	1	0.6	0.8832 *(0.1140)*
1	3	2	0	0.0	0.6401 *(0.1806)*
1	3	3	0	0.0	0.2664 (0.1490)
2	1	1	5	2.8	0.9925 *(0.0076)*
2	1	2	7	3.9	0.9617 *(0.0312)*
2	1	3	3	1.7	0.8063 *(0.1318)*
2	2	1	13	7.2	0.9142 *(0.0590)*
2	2	2	9	5.0	0.6852 *(0.1353)*
2	2	3	11	6.1	0.2955 *(0.1529)*
2	3	1	7	3.9	0.5776 *(0.2388)*
2	3	2	5	2.8	0.2458 *(0.1670)*
2	3	3	7	3.9	0.0611 *(0.0674)*
3	1	1	0	0.0	0.7249 *(0.1588)*
3	1	2	0	0.0	0.3734 *(0.1872)*
3	1	3	0	0.0	0.1069 *(0.0916)*
3	2	1	3	1.7	0.2292 *(0.1752)*
3	2	2	8	4.4	0.0559 *(0.0574)*
3	2	3	19	10.6	0.0101 *(0.0122)*
3	3	1	3	1.7	0.0501 *(0.0722)*
3	3	2	7	3.9	0.0082 *(0.0117)*
3	3	3	33	18.3	0.0013 *(0.0019)*
TOTAL			180	100.0	

[a] 1, 2, and 3 denote expert classifications into the intervals > 90%, 50–90%, and < 50% probabilities of asthma, respectively.

The diagnostic accuracies of the three experts were unknown and so all three opinions were combined into a single probability label for each of the reviewed individuals.

DISCUSSION

The aim of this project was to use neural networks to rank a population in order of individual probability of asthma, based on their responses to a screening questionnaire. Typically, network-training examples would be labelled with a known outcome, for example the presence or absence of asthma. One of the problems in providing such labels is the inherent uncertainty in the diagnosis of asthma, as reflected in the lack of agreement between the three experts. Similar disagreement has occurred in similar circumstances in another study (4) where even permitting the experts to discuss problem cases did not result in complete agreement. However, the varying opinions of the experts were a major strength in the neural network method because the cases where the experts disagreed held as much information as those where they agreed. The 27 varying permutations of expert opinion allowed the Bayesian model to capture the uncertainty for each patient in a single probability label.

The development of the model can be viewed as an example of latent class analysis (27, 28). The parameters for each expert were evaluated independently and no assumptions were made regarding the competence of each expert or their interpretation of the diagnostic categories. The 13 model parameters are analogous to the terms 'sensitivity' and 'specificity'. The three experts had no contact with each other and no knowledge of the identity of the respondents or their questionnaire responses. The resulting labels therefore provided a 'gold-standard' that was external to the questionnaire even though there was inevitably some overlap in the information requested from patients in the questionnaire and that collected at clinical assessment.

TABLE 2. Expected asthmatics in sets selected for review above various screening thresholds when using the neural network to rank the population (n = 6825)

Screening threshold[a]		Number of individuals to review (no asthma diagnosis)	Number and percent of reviewed individuals expected asthmatic	
Percent	*Number*		Number	*Percent*
1	68	12	11.10	93%
5	341	68	60.14	88%
10	683	239	177.60	*74%*
15	1024	474	304.29	*64%*
20	1365	746	398.24	*53%*

[a] Cut-off points when population (n = 6825) ranked according to probability of asthma (for example, top 1% of population, top 5% of population, and so on).

Of particular importance are the hitherto unrecognised asthmatics. In order to assess the scale of this under-diagnosis problem it would be helpful to have an estimate of the number of asthmatics in the whole community and the proportion of these unknown to the medical services. The neural network system predicts a probability of asthma for each individual in the population. The expected number of asthmatics is calculated by summing the probabilities associated with each individual. The extreme probability values of 0 and 1 representing 100% certainty of absence or presence of asthma respectively were absent because the experts were not allowed to express this level of certainty in their diagnoses, the most extreme intervals being $< 50\%$ and $> 90\%$ probability of asthma, respectively. Any attempt to estimate prevalence over the whole population is therefore necessarily inflated due to the considerable contribution to the prevalence sum by the very large number of non-asthmatics, each contributing a small non-zero amount. This model is, therefore, not suitable for estimating the prevalence of asthma in the entire population. However, a crude method based on clinical intuition and the number of positive responses estimated that the prevalence of asthma in this survey population was approximately 13.8%, of whom 6.6% were unknown to the medical services (7).

The actual number of asthmatics identified would depend on how far down the ranked list the screening cut-off was set, as shown in Table 2. Of more interest in the context of limited resources are the expected numbers of asthmatics above various percentage thresholds and not already known to the medical services, giving an indication of the false positive rate when using that threshold as a screening cut-off. This measure is comparable with the positive predictive value of conventional screening tests. The threshold used in any particular community would need to be set according to the resources available for the follow up clinical review.

Since these estimates of expected numbers of asthma diagnoses depend on the fact that the training set had been randomly selected from within population strata, it is important to be aware of possible sample bias. First, there is the possibility of bias due to non-response in the postal survey. Random 5% samples from the responders and non-responders groups were compared with respect to age, gender and medical practice records in the last year of asthma diagnosis, prescription for bronchodilators, number of respiratory consultations and total number of consultations. No material difference was found between the two groups. Second, there is the possibility of bias due to the exclusion of incomplete questionnaires. For both training set and wider population the distribution of numbers of positive responses before exclusion was compared with the distribution after exclusion. Chi-squared goodness of fit tests revealed no significant difference. Third is the possibility of bias due to those individuals who were selected but did not attend for clinical review. Analysis of the reasons for non-attendance revealed that almost all were because the general practitioner had been unable to make contact to issue the invitation, a reason not directly related to probability of asthma. In addition, examination of the variables of age, gender, asthma diagnosis, and prescriptions for bronchodilators or steroids in the last year revealed no important difference between the selected sample and those who attended for clinical review. Analyses had therefore revealed no evidence of sample bias from these three sources.

When using sophisticated methods, it is important to bear in mind cost-benefit issues. A simple scoring system is currently used in the WYCAP study to screen for possible asthmatics based on the number of positive responses to six key questions (7). Ranking using this scoring system does not achieve the individual prioritisation provided by more sophisticated methods. Two other sophisticated methods, namely logistic regression and a probabilistic neural network (Bayesian classifier) were tested. There was no difference in ranking accuracy between logistic regression and the neural network, but both fared better than the probabilistic network. The primary purpose of this work was to develop a methodology, which could be applied generally to rank patients according to diseases which are difficult to diagnose. The advantages of the neural network model include not only its flexibility in identifying the best model given the data, but also its capacity for incorporating all available information including prior probabilities.

The experts diagnosed one of three categories. The Bayesian methods and Monte-Carlo techniques transformed these into 27 possible labels for the training questionnaires. The neural network training produced a regression model, which was able to generalise and predict on an individual basis the probability of asthma of questionnaire respondents from the wider community. The estimation of expected numbers of asthmatics gave an indication of the effectiveness of various screening cut-offs in detecting asthmatics in the community. In conclusion, this study has shown that it is possible, by combining a variety of statistical techniques, to use a simple questionnaire to rank a population in order of probability of asthma, providing a method of finding asthmatics in the community. We now plan to validate our methodology by detailed clinical assessment of a random sample of patients from our most recent survey conducted in 2001. Asthmatic status as determined clinically will be compared with the ranking position predicted by the system.

REFERENCES

1. Weiss KB. An overview of recent trends in asthma epidemiology. Eur Resp Rev. 1996;6(35):101–104.
2. Omran M, Russell G. Continuing increase in respiratory symptoms and atopy in Aberdeen schoolchildren. BMJ. 1996;312:34.

3. Kaur B, Anderson HR, Austin J, Burr M, Harkins LS, Strachan DP, et al. Prevalence of asthma symptoms, diagnosis and treatment in 12 to 14 year old children across Great Britain. International study of asthma and allergies in childhood (ISAAC UK). BMJ. 1998;316:118–124.
4. De Marco R, Cerven I, Bugiani M, Ferrari M, Berlato G. An undetected burden of asthma in Italy: The relationship between clinical and epidemiological diagnosis of asthma. Eur Resp J. 1998;11:599–605.
5. Speight ANP, Lee DA, Hey EN. Underdiagnosis and undertreatment of asthma in childhood. BMJ. 1983;286:1253–1256.
6. Pederson S. Early use of inhaled steroids in children with asthma. Clin Exp Allergy. 1997;27:995–1006.
7. Frank PI, Ferry S, Moorhead T, Hannaford P. Use of a postal questionnaire to estimate the likely under-diagnosis of asthma-like illness in adults. Br J Gen Pract. 1996;46:295–297.
8. Frank TL, Frank PI, McNamee R, Wright T, Hannaford P, Morrison J, et al. Assessment of a simple scoring system applied to a screening questionnaire for asthma in children aged 5–15 years. Eur Resp J. 1999;14(5):1190–1197.
9. Cross S, Harrison RF, Kennedy RL. Introduction to neural networks. Lancet. 1995;346:1075–1079.
10. Tarassenko L. A Guide to Neural Computing Applications. London: Arnold; 1998:12–18.
11. Hirsch S, Turega M, Frank P. A feasibility study: The use of an artificial neural network in detecting asthma from a medical screening questionnaire. In: Ifeachor EC, ed. Proceedings of the International Conference on Neural Networks and Expert Systems in Medicine and Healthcare. Plymouth, UK: University of Plymouth; 1996:36–42.
12. Hirsch S, Frank PI, Shapiro J. Use of an artificial neural network in estimating prevalence and assessing under-diagnosis of asthma. Neural Comput Appl. 1997;5(2):124–128.
13. Beale R, Jackson T. Neural Computing. An Introduction. London: Institute of Physics Publishing; 1992;63–94.
14. Bishop CM. Neural Networks for Pattern Recognition. Oxford: Clarendon Press; 1995:116–163.
15. Patterson DW. Artificial Neural Networks, Theory and Applications. Singapore: Prentice Hall; 1996:141–179.
16. Burney PGI, Luczynska C, Chinn S, Jarvis D. The European community respiratory health survey. Eur Resp J. 1994;7:954–960.
17. Hirsch S. The Assessment of Artificial Neural Networks and Other Statistical Techniques in Asthma Screening. Ph.D. Thesis. University of Manchester, England; 1998.
18. Lilford RJ, Braunholtz D. The statistical basis of public policy: A paradigm shift is overdue. BMJ. 1996;313:603–607.
19. Metropolis NAW, Rosenbluth MN, Rosenbluth AH, Teller AH, Teller E. Equations of state calculations by fast computing machines. J Chem Physics. 1953;21:1087–1091.
20. Neal RM. Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical Report CRG-TR-93–1. University of Toronto, Canada, Department of Computer Science; 1993.
21. Krauth W. Introduction To Monte Carlo Algorithms. Transcript of lectures given in '96 summer schools in Beg-Rohu (France) and Budapest. CNRS-Laboratoire de Physique Statistique, Ecole Normale Superieure. Paris; 1997.
22. Tarassenko L. A Guide to Neural Computing Applications. London: Arnold; 1998:112.
23. Daly F, Hand DJ, Jones MC, Lunn AD, McConway KJ. Elements of Statistics. Wokingham, England: Addison-Wesley Publishing Company; 1995:442–443.
24. Tarassenko L. A Guide to Neural Computing Applications. London: Arnold; 1998:71.
25. Pickles A, Dunn G. Screening for stratification in two-phase ('two stage') epidemiological surveys. Stat Meth Med Res. 1995;4:73–89.
26. Norusis MJ. SPSS for Windows Base System User's Guide, Release 6.0. Chicago, IL: SPSS Inc.; 1993:215.
27. Formann A, Kohlmann T. Latent class analysis in medical research. Stat Meth Med Res. 1996;5:179–211.
28. Uebersax JS, Grove W. Latent class analysis of diagnostic agreement. Stat Med. 1990;9:559–572.

Appendix 1

The Labelling Model Parameters

(Notation as described in text)

The 13 independent parameters in the model were:

$\{p_0, p_{A1}, p_{A2}, p_{B1}, p_{B2}, p_{C1}, p_{C2}, q_{A1}, q_{A2}, q_{B1}, q_{B2}, q_{C1}, q_{C2}\}$.

The 20 parameters could then be expressed as:

$\{p_0, 1 - p_0, p_{A1}, p_{A2}, 1 - p_{A1} - p_{A2}, \ \ldots\ q_{C1}, q_{C2}, 1 - q_{C1} - q_{C2}\}$.

Appendix 2

The Labelling Model Prior and Likelihood Distributions

(Notation as described in text)

The Prior

Condition 1: 'Expert is better than a guesser' requires for each expert i:

$$q_{i1}(1 - p_0) + q_{i2}(1 - p_0) + p_{i3}p_0 < \min[1 - p_0, p_0]$$

Condition 2: 'Expert is informative' requires for each expert i:

$$p_{i1} + p_{i2} > q_{i1} + q_{i2}.$$

The Likelihood

The probability of any given data point R, S, T is:

$$P(R, S, T \mid parameters) = p_0 p_{AR} p_{BS} p_{CT} + (1 - p_0) q_{AR} q_{BS} q_{CT}.$$

Assuming that the 180 data points are independent the probability of the whole set of data, $P(data \mid parameters)$, is the product of the probabilities of each data point, that is:

$$p(N_{111}, N_{112}, \ldots \mid p_0, p_{A1}, p_{A2}, \ldots) =$$
$$[p_0 p_{A1} p_{B1} p_{C1} + (1 - p_0) q_{A1} q_{B1} q_{C1}]^{N_{111}}$$
$$[p_0 p_{A1} p_{B1} p_{C2} + (1 - p_0) q_{A1} q_{B1} q_{C2}]^{N_{112}}$$
$$[p_0 p_{A1} p_{B1} p_{C3} + (1 - p_0) q_{A1} q_{B1} q_{C3}]^{N_{113}}$$
$$\ldots$$
$$[p_0 p_{A3} p_{B3} p_{C3} + (1 - p_0) q_{A3} q_{B3} q_{C3}]^{N_{333}}$$