Thorax 1989;**44**:1031–1035

Prediction of microbial aetiology at admission to hospital for pneumonia from the presenting clinical features

Prepared on behalf of the British Thoracic Society Pneumonia Research Subcommittee by
B M FARR, D L KAISER, B D W HARRISON, C K CONNOLLY

ABSTRACT The selection of initial antimicrobial treatment in a patient with community acquired pneumonia is an important clinical decision. Because this decision is usually made before the results of specific microbiological tests are available, we attempted to determine how well the presenting clinical features would allow prediction of microbial aetiology in 441 adults admitted to hospital with pneumonia. Five of 90 variables available on admission were selected for inclusion in a multivariate discriminant function analysis because of their strong association with one or more of the major aetiological subsets (*Mycoplasma pneumoniae, Streptococcus pneumoniae*, "other," and undetermined). These variables were age, number of days ill before admission, presence or absence of bloody sputum and of lobar infiltration on chest radiograph, and white blood cell count. The microbial aetiology was correctly predicted by this discriminant function analysis in only 42% of cases, which gives a quantitative estimate of the degree of difficulty encountered in determining the microbial aetiology at the time of admission for pneumonia. When a similar discriminant function analysis was applied to the third of patients in whom the microbial aetiology was never determined, most of these cases were predicted to be due to *Streptococcus pneumoniae*.

Introduction

The specific microbial diagnosis in cases of community acquired pneumonia is often delayed for two or more days until cultures of sputum or blood become positive. Culture techniques are not available for some pathogens and positive serological results are not obtained in some cases for several weeks.

Knowledge of microbial aetiology would be of value at the time of hospital admission, allowing selection of appropriate antibiotics. The sputum Gram stain may be helpful for rapid diagnosis of a bacterial infection. Finding Gram positive, lancet shaped diplococci in sputum suggests pneumococcal pneumonia, for which the Gram stain is specific but relatively insensitive.[1 2] Gram stain diagnosis of other microbes in sputum has not been subjected to such formal evaluation.

Sputum is not available in about one third of patients on admission[1] and when it is available the Gram stain may not provide a clear indication of microbial aetiology. Other rapid tests, such as sputum counterimmunoelectrophoresis for pneumococcal antigen and the serum immunofluorescence antibody test for mycoplasma specific IgM, may be helpful but are not widely available. For these reasons we have undertaken an analysis of the possibility of predicting microbial aetiology given the presenting clinical features at the time of hospital admission.

In up to a third of patients admitted to hospital with community acquired pneumonia the microbial aetiology is never determined. Possibly some of these infections are due to currently unknown pathogens. To understand the aetiology of these undetermined cases better, we have compared the presenting clinical features of these patients with those of patients with an identified microbial aetiology. On the assumption that some of the unidentified cases might be due to recognised microbial pathogens, a discriminant function analysis was used to show the aetiological

This study was organised by the British Thoracic Society pneumonia research subcommittee and the Public Health Laboratory Service. Members of the subcommittee were B E Andrews, C L R Bartlett (epidemiological coordinator), C K Connolly (chairman), D A Ellis, B M Farr, C D Gooch, B D W Harrison (clinical coordinator), J T MacFarlane, A D Macrae, D Miller, S Pugh, and J B Selkon.

Address for reprint requests: Dr Barry M Farr, Department of Medicine, Box 473, University of Virginia Medical Center, Charlottesville, Virginia 22908, USA.

Accepted 3 August 1989

category into which the undetermined cases would best fit.

Methods

This analysis was performed on data from 453 patients enrolled in the British Thoracic Society pneumonia study at the time of their admission to 25 hospitals throughout Great Britain during 1982 and 1983. The methods for this study have been described.[1]

All adults aged 15–74 years admitted under the care of the participating physicians with community acquired, primary pneumonia were enrolled in the study. Data were collected on specially designed forms to document the history, results of physical examination and laboratory investigations, and treatment.

MICROBIOLOGICAL METHODS

Sputum, blood and urine specimens were obtained under routine clinical conditions and were submitted to the local microbiology laboratory, where culture and Gram staining of sputum, blood cultures, and complement fixation tests for psittacosis, Q fever, influenza A and B, respiratory syncytial virus infection, and parainfluenza were performed. All laboratories participated in the Public Health Laboratory Service national quality control scheme. From these specimens aliquot samples were forwarded to the Public Health Laboratory at Nottingham for serological testing for *Legionella pneumophila*[3] and countercurrent immunoelectrophoresis for pneumococcal antigen[4] and to the Mycoplasma Reference Laboratory in Norwich for *Mycoplasma pneumoniae* complement fixation testing and for mycoplasma specific IgM and IgG immunofluorescence antibody testing.[5]

CRITERIA FOR LABORATORY DIAGNOSIS OF INFECTION

Pneumococcal infection was defined for the purposes of this analysis by isolation of *Streptococcus pneumoniae* from sputum, blood, or pleural fluid, or by detection of pneumococcal antigen in sputum, serum, or urine by countercurrent immunoelectrophoresis. *Haemophilus influenzae* and *Staphylococcus aureus* infections were defined by isolation of these organisms from sputum, but for other bacteria "infection" required a report that the organism was predominant in the sputum Gram stain in addition to being isolated from sputum. Criteria for serological diagnosis in this study have been described previously.[1]

DATA HANDLING AND STATISTICAL METHODS

Ninety independent variables, obtained from the history and from the results of physical examination, laboratory tests, and initial antibiotic treatment, were examined for association with microbial aetiology.

For the purposes of this analysis 12 patients were excluded because of dual infection with at least two of the three main categories of identified micro-organisms—pneumococcus, mycoplasma, and "other." Microbes identified in the 78 patients in the category "other" are listed in table 1. The observations on the remaining 441 patients were used to study the relation of presenting clinical features to the four main aetiological categories—*Streptococcus pneumoniae*, *Mycoplasma pneumoniae*, "other," and "undetermined."

Data were coded manually before being entered into the computer and verified at the Public Health Laboratory Service Communicable Disease Surveillance Centre. Range and consistency checks were performed on all data before analysis. Frequency distributions were examined for each variable, and logarithmic transformations were performed as necessary.

Univariate statistical analysis

For univariate analysis cross tabulation and χ^2 testing were used for categorical variables, and Fisher's exact test when expected values were small. Continuous variables were analysed by Student's t test, analysis of variance, or linear regression. Continuous variables with obvious departure from a normal distribution were analysed by interval grouping and contingency table χ^2 tests. This work was performed with a Minitab statistical package on a PDP-11 computer.[6]

Multivariate statistical analyses

Independent variables associated with causative organisms in the univariate analysis at the 5% level of significance were used in a multivariate discriminant

Table 1 *Microbes identified in 78 patients in the category "other"**

Organism	*Frequency*
Bacteria:	
Haemophilus influenzae	18
Chlamydia psittaci	10
Legionella pneumophila	7
Coxiella burnetii	5
Staphylococcus aureus	4
Enterobacteriaceae	4
Streptococcus spp (2 group A, 2 group C)	4
Bacteroides spp	1
Haemophilus haemolyticus	1
Streptococcus milleri	1
Viruses:	
Influenza A	24
Respiratory syncytial virus	2
Varicella	2
Adenovirus	1

*Total number of microbes identified greater than 78 because 13 patients had more than one organism.

function analysis. The analysis was run stepwise, including predictors in the order of their statistical importance, with a 0·05 inclusion criterion. This analysis resulted in an equation that could be used to predict the aetiological category for individual patients.

A second discriminant function analysis was then used to predict into which of the three categories of identified cause the "undetermined" cases would best fit. This was done by first performing a stepwise discriminant function analysis in which we used only the three categories of identified cause (that is, with the group of "undetermined" aetiology excluded). The equation generated from this analysis was then used to predict which of the three categories of identified cause the "undetermined" cases would best fit.

A final analysis was performed to predict once again which of the categories of identified cause the "undetermined" cases would best fit; this time only the pneumococcal and mycoplasmal categories were used in the preliminary stepwise discriminant function analysis because these two were the most frequent and homogeneous categories.

All multivariate analyses were performed with the SAS statistical package.[7]

Results

Presenting clinical features which showed a significant association with cause in the univariate analyses are shown in table 2.

Patients with *M pneumoniae* pneumonia tended to be two decades younger; to have been ill longer before admission; and to have less leucocytosis and lower frequencies of chest pain, of lobar infiltrate, and of bloody sputum. Contrary to the classical descriptions of dry cough with mycoplasma infection in outpatients, sputum production was reported in about the same proportion of each of the four main aetiological categories in these patients admitted to hospital. Although most patients with mycoplasma infection were young adults, five (7%) were over 50, and one was 72; *M pneumoniae* accounted for one third of the cases of pneumonia among adults below the age of 50 years. Patients with *M pneumoniae* were more likely to have received an antibiotic before admission (85%) than patients with non-mycoplasmal pneumonia (38%); they mainly had antibiotics that were not effective against mycoplasma.

Patients with pneumococcal pneumonia were similar in age and sex distribution to the patients in the aetiological categories "other" and "undetermined." The white blood cell count was significantly higher than in patients in the other three categories, though three patients (2·4%) had leucopenia. Pneumococcal pneumonia tended to be associated with bloody sputum more frequently than did cases in the mycoplasmal and "undetermined" categories. There was no significant difference in the proportion with herpes between the pneumococcal (12%) and the non-pneumococcal patients (7%) despite the large sample size. The positive predictive value of herpes labialis for *S pneumoniae* was only 45% in this study, suggesting that this may not be a reliable sign of pneumococcal infection itself but be a non-specific phenomenon in febrile patients with pneumonia.

The presenting clinical features of patients with pneumococcal pneumonia most closely resembled those in cases of "undetermined" aetiology. These two groups were similar in age, number of days ill before admission, the proportion with chest pain, and the proportion with lobar infiltrate. The major difference between patients with pneumococcal and undetermined pneumonia was in the frequency of antibiotic treatment before admission.

Patients with "other" types of pneumonia also resembled pneumococcal pneumonia patients in age and number of days of illness before admission, but tended to have less leucocytosis; a lower proportion had chest pain and lobar infiltrates and a higher proportion had received antibiotics before admission.

The first multivariate discriminant function analysis

Table 2 *Presenting clinical features of major aetiological categories*

	Undetermined (n = 148)	*Pneumococcal (n = 142)*	*Mycoplasmal (n = 73)*	*Other (n = 78)*
Age (mean (SEM), years)	48·1 (1·5)	54·2 (1·4)	33·7 (1·6)	53·5 (1·8)***
Number of days ill before admission (mean (SEM))	6·6 (0·92)	6·0 (0·84)	10·6 (0·72)	7·1 (0·98)***
White blood count (mean (SEM), × 10^9/l)	13·6 (0·56)	15·1 (0·62)	11·2 (0·64)	12·6 (0·70)***
Number (%) of patients having:				
antibiotics before admission	67 (45)	32 (22)	61 (84)	37 (47)***
chest pain	107 (72)	103 (72)	33 (45)	40 (51)***
lobar infiltrate	69 (47)	74 (52)	20 (27)	26 (33)**
bloody sputum	15 (10)	32 (22)	1 (1)	11 (14)**

**$p < 0·01$; ***$p < 0·001$ (contingency table χ^2 comparing the frequency of a particular variable in the four groups or analysis of variance comparing the mean values for the four groups).

Table 3 *Independent variables contributing to the multivariate discriminant function analysis*

Variable	*Partial R^2*	*F statistic*	*p value**
Age > 40 years	0·2871	52·555	0·0001
WBC $\geq 14{\cdot}4 \times 10^9/l$	0·0923	13·225	0·0001
No of days before admission ≥ 9	0·0680	9·449	0·0001
Lobar infiltrate	0·0554	7·559	0·0006
Bloody sputum	0·0311	4·127	0·0172

*p value refers to the probability associated with a variable after multivariate adjustment for all other variables listed in the table.

resulted in the inclusion of five independent variables in the predictive model: age, duration of illness before admission, white blood cell count, and presence or absence of lobar infiltrate and bloody sputum (table 3). With this model 167 of 396 (42%) patients' aetiological categories were correctly predicted (table 4). Of the four major categories, mycoplasmal infections were most frequently predicted correctly (77%). Pneumococcal and "other" infections were correctly predicted only about half the time, and pneumonia of undetermined cause was poorly distinguished from the other categories.

The results of the second and third discriminant function analyses regarding the aetiology of the "undetermined" cases of pneumonia are shown in table 5. When the discriminant function analysis was based on three categories of identified cause most of the cases of unidentified cause were predicted to be pneumococcal. When the analysis was based on only the two most frequent and homogeneous aetiological categories (*S pneumoniae* and *M pneumoniae*), an even larger majority of cases of unidentified cause were predicted to be pneumococcal.

Discussion

A physician evaluates many factors in the history and physical findings of a patient with pneumonia before prescribing an antimicrobial drug for the pathogen considered to be the most likely aetiological agent. Such a decision is often made before the results of specific microbiological tests are available. The results of this study provide a quantitative assessment of the difficulty encountered in the differential diagnosis of the infecting organism at the time a patient is admitted to hospital with community acquired pneumonia.

The results of our multivariate discriminant function analysis suggest that among patients presenting with community acquired pneumonia patients with *Mycoplasma pneumoniae* are easiest to distinguish from the rest, but even with a multivariate equation 23% of patients with mycoplasma infection were inaccurately assigned to another aetiological category.

The poor precision of our multivariate discriminant function in differentiating pneumococcal and "undetermined" pneumonia may be due to partial overlap of these categories, some patients with true pneumococcal infection being included in the "undetermined" category, and perhaps to uncertainty introduced by the wide range of "other" pathogens included in the category "other." When only the three major categories of identified cause were used the identification of cases of pneumococcal pneumonia was improved, and most of the "undetermined" cases were predicted to be pneumococcal. This accords with the results of previous studies, which have suggested that an appreciable proportion of cases of undetermined aetiology may be due to *S pneumoniae*.[8,9] Antibiotic treatment before admission has a dramatic effect on the viability of pneumococci and patients receiving antibiotics before admission have significantly lower rates of culture of pneumococci than patients not receiving antibiotics before admission.[1] Another reason why patients may fall into the undetermined category is incomplete diagnostic testing, which is particularly important for pneumococcal pneumonia.[1] Most of the tests for *S pneumoniae* are rather insensitive, and the most sensitive pneumococcal test (sputum countercurrent immunoelectrophoresis) is the least frequently performed test for *S pneumoniae*.[1] Possibly some of the undetermined pneumonias are not infections but the inflammatory response to aspiration of acidic gastric contents (Mendelson's syndrome).[10]

Previous studies have suggested that it is difficult to predict microbial aetiology on the basis of presenting clinical features[11] and we have provided a quantitative estimate of this difficulty. We found that the four main categories of microbial causes of pneumonia could be correctly predicted by using a computerised discrimin-

Table 4 *Prediction of aetiological category using all cases to create the discriminant function*

Actual aetiological category	*Predicted category (No (%))*			
	Mycoplasmal	*Other*	*Pneumococcal*	*Undetermined*
Mycoplasmal	51 (77·3)	10 (15·2)	1 (1·5)	4 (6·1)
Other	12 (16·4)	39 (53·4)	11 (15·1)	11 (15·1)
Pneumococcal	12 (9·6)	36 (28·8)	56 (44·8)	21 (16·8)
Undetermined	30 (22·7)	39 (29·6)	42 (31·8)	21 (15·9)

Table 5 *Prediction of aetiological category for undetermined cases using only cases with identified cause to create the discriminant function*

No of aetiological categories used in creating discriminant function	*Predicted category (No (%))*		
	Pneumococcal	*Mycoplasmal*	*Other*
3*	76 (57·6)	31 (23·5)	25 (18·9)
2†	101 (76·5)	31 (23·5)	—

*The three categories were *Mycoplasma pneumoniae*, *Streptococcus pneumoniae*, and "other."
†The two categories were *M pneumoniae* and *S pneumoniae*.

ant function analysis only 42% of the time on the basis of presenting symptoms and signs at admission to hospital; this analysis allowed prediction of a particular species (*S pneumoniae* or *M pneumoniae*) in 107 of 441 cases (24·3%). Proper use of the microbiology laboratory should increase the precision of the initial clinical diagnosis by use of rapid diagnostic tests such as the sputum Gram stain, sputum countercurrent immunoelectrophoresis for pneumococcal antigen, and a serum immunofluorescence antibody test for IgM antibody to *M pneumoniae*.[1]

The research committee of the British Thoracic Society acknowledges with gratitude the cooperation of the local coordinating physicians: Dr M K Benson (Oxford), Dr C K Connolly (Northallerton and Darlington), Dr D Eraut (Southend), Dr I W Fawcet (Yeovil), Dr B D W Harrison (Norwich), Dr T Higenbottam (Papworth), Dr P B Iles (Birmingham), Dr P F Jenkins (Norwich), Dr O G Jones (Hull), Dr D H Kennedy (Glasgow), Dr J Maberly (Keighley), Dr J T Macfarlane (Nottingham), Dr M Muers (Leeds), Dr I C Paterson (Lincoln), Dr J M Shneerson (Newmarket), Dr D Sinclair (Middlesborough), Dr J E Stark (Papworth), Dr T C Stokes (Woolwich), Dr G O Thomas (Abergavenny), Dr A K Webb (Manchester), Dr R J White (Bristol), our consultant physician colleagues who also entered patients, their consultant microbiology colleagues, their junior staff, and their secretaries.

They would also like to thank Miss J White and Mrs M Ho of the Communicable Diseases Surveillance Centre, for data entry and verification, Dr G Manning of the Communicable Diseases Surveillance Centre for assistance with computer programming and Mrs M Sillis for performing all the mycoplasma serological tests, Mr A Laverick for performing the legionella serological tests and Mr R Pilkington for performing the pneumococcal countercurrent immunoelectrophoresis tests, and Mrs M Ward for typing the manuscript.

References

1 British Thoracic Society Research Committee. Community-acquired pneumonia in adults in British hospitals in 1982–1983: a BTS/PHLS survey of aetiology, mortality, prognostic factors and outcome. *Q J Med* 1987;NS**62**:195–220.
2 Rein MF, Gwaltney JM Jr, O'Brien WM, Jennings RH, Mandell GL. Accuracy of Gram's stain in identifying pneumococci in sputum. *JAMA* 1978;**239**:2671–3.
3 Macfarlane JT, Finch RG, Laverick A, Macreae AD. Pittsburgh pneumonia agent and Legionellosis in Nottingham. *Br Med J* 1981;**283**:1222.
4 El-Rafaie M, Dulake C. Counter-current immunoelectrophoresis for the diagnosis of pneumococcal chest infections. *J Clin Pathol* 1975;**28**:801–6.
5 Wreghitt TG, Sillis M. A μ capture ELISA for detecting *Mycoplasma pneumoniae* IgM: comparison with indirect immunofluorescence and indirect ELISA. *J Hyg Camb* 1985;**94**:217–27.
6 Ryan TA, Joiner BL, Ryan BF. *Minitab reference manual*. University Park, Pennsylvania: Pennsylvania State University Press, 1982.
7 SAS Institute. *SAS user's guide: statistics*. Version 5. Cary, North Carolina: SAS Institute, 1985.
8 Bath JCJL, Boissard GPB, Calder MA, Moffat MAJ. Pneumonia in hospital practice in Edinburgh 1960–1962. *Br J Dis Chest* 1964;**58**:1–16.
9 Spencer RC, Philp JR. Effect of previous antimicrobial therapy on bacteriological findings in patients with primary pneumonia. *Lancet* 1973;ii:349–51.
10 Bartlett JG, Gorbach SL. The triple threat of aspiration pneumonia. *Chest* 1975;**68**:560–6.
11 Woodhead MA, Macfarlane JT. Comparative clinical and laboratory features of legionella with pneumococcal and mycoplasma pneumonias. *Br J Dis Chest* 1987;**81**:133–9.