Prediction of chronic neonatal lung disease in very low birthweight neonates using clinical and radiological variables

Steven W Ryan, Nicholas J Wild, Rosemary J Arthur, Ben N J Shaw

Abstract

There are good theoretical reasons for earlier intervention in neonates likely to develop chronic neonatal lung disease (CNLD). Very low birthweight (VLBW) neonates who receive artificial ventilation are at high risk of CNLD. A test was therefore developed to predict CNLD based on clinical and radiological information readily available at 7 days of age in VLBW neonates. Logistic regression analysis was used to identify those factors significantly and independently associated with CNLD.

For each neonate it was possible to insert the value of the independent factors into the equation, providing a probability value between 0 and 1. By selecting different cut off values between 0 and 1, and knowing which neonates had developed CNLD, it was possible to assess the use of varying probability values as a predictive test for CNLD. The variation in these two parameters was graphically represented by a receiver operator characteristic (ROC) curve.

The area under the ROC curve was used to represent the discriminatory capacity of the test over its full range of values. The maximum area under an ROC curve is unity. The area under the ROC curve was similar in a model with and without radiographic information (0.926 and 0.913 respectively) and was 0.937 in neonates from another hospital.

(Arch Dis Child 1994; 71: F36–F39)

Chronic neonatal lung disease (CNLD) is likely to remain a serious problem despite the introduction of surfactant treatment and the use of antenatal corticosteroids. Although some studies have shown a reduction in the frequency of CNLD when using surfactant treatment,1 CNLD still develops in many very low birthweight (VLBW) neonates.2 There was an increase in the frequency of CNLD during the period when surfactant treatment became available,3 partly due to the increasing survival of neonates below 28 weeks’ gestation.

In the past interventions in CNLD have been undertaken when the disorder was established.4 It is now suggested that earlier ‘preventative’ treatment would be more appropriate, allowing intervention in the dynamic process of abnormal healing and continuing lung damage occurring during the first postnatal week.5 Potential interventions include suppressing the inflammatory response that causes abnormal healing or a systematic attempt to reduce ventilator induced barotrauma. Inflammation could be suppressed using dexamethasone, which has several potentially serious side effects.6 A reduction in barotrauma could result from increased tolerance of respiratory acidosis, which is also associated with an increased risk of complications.7

If such interventions are contemplated as part of routine care or research, it is important that they are limited, as far as possible, to neonates at high risk of CNLD.

Many factors (clinical, radiological, immunological, and biochemical) have been associated with CNLD, helping us to understand its pathogenesis, but by themselves none provide a strong predictive test. We wished to develop a simple clinical and radiological test for CNLD based on known or demonstrable associations in (VLBW) neonates who are at high risk of CNLD. To be widely applicable, easily identifiable, and measurable clinical factors would be needed. A test undertaken at 7 days of age would allow the features of abnormal healing, inherent in CNLD, to develop.

Factors associated with CNLD can be conveniently divided into four groups. The first group relates to birth parameters: young gestational age and male sex. The second group relates to the severity of hyaline membrane disease. Severity can be gauged by radiographic appearances and ventilator settings during the first 48 hours, providing the latter are adjusted in a standard way. The third group relates to failed recovery from hyaline membrane disease, occurring from day 2. Abnormal healing should manifest as a delay in improvement in ventilator status towards the end of the first week. Patent ductus arteriosus is also often seen during this period, and is associated with CNLD.8 The fourth group relates to abnormal radiological gas patterns. Neonates developing air leaks are at greater risk of developing CNLD.9 Terminal airway distension may also represent a more minor degree of lung barotrauma.10

Factors from all four of these groups may be used to predict CNLD in VLBW neonates. Because of the possibility of interassociation between some of these explanatory factors, it would be necessary to use a multiple regression technique to attribute association independently. These associations would be the basis of a predictive test.
Patients and methods

Patients

VLBW neonates born in the calendar years 1988–9, and admitted to the neonatal units at Leeds General Infirmary or Liverpool Maternity Hospital for ventilation, were identified. Data from neonates in Leeds were used to develop a predictive test. Data from Liverpool neonates were used to assess its reproducibility. There were 166 neonates from Leeds and 133 from Liverpool who fulfilled the above criteria. Their case records and chest radiographs were reviewed. Table 1 gives the basic clinical parameters for the two groups.

Methods

The birth variables gestation (weeks) and sex were recorded. During the first 48 hours the following variables were recorded: radiographic grade of hyaline membrane disease (grades 1 to 4); maximum peak inspiratory pressure (cm H₂O); maximum inspired oxygen concentration (%); and maximum ventilator rate (breaths/minute). Abnormal gas patterns, air leaks (pneumothoraces, pneumomediastinum, or pulmonic emphysema), and terminal airway distension were also recorded blindly by one investigator (RJA). These data were coded as categorical variables. When there was no abnormality it was coded as 0 and any abnormal gas pattern was coded as 1.

During the next five day period the following variables were recorded: maximum peak inspiratory pressure; maximum inspired oxygen concentration; and maximum ventilator rate. The presence of a patent ductus arteriosus was recorded on clinical grounds. This required the presence of a continuous or systolic murmur, features consistent with heart failure, including pulmonary plethora, which led to a change in medical management. Ventilation status at the end of seven days was also recorded.

Neonates were categorised as having CNLD at 28 days of age if they remained dependent on supplemental oxygen and they had an abnormal chest radiograph.11

The above explanatory variables were entered into a logistic regression program (SPSS Release 4.0, SPSS Inc, Chicago) to determine the independent degree of association with CNLD. Factors that were non-significant were discarded, so that a final model with the least number of explanatory variables was made.

For each neonate it was possible to insert the value of the independent factors into the equation, providing a probability value between 0 and 1. By selecting different cutoff values between 0 and 1, and knowing which neonates had developed CNLD, it was possible to assess the use of varying probability values as a predictive test for CNLD. For each group, for example, the Leeds group, the process results in 166 probability values between 0 and 1. By selecting a single cut off value, for example 0.5, it is possible to divide the neonates into two groups, one with probability values <0.5 and one with probability values ≥0.5. Knowing the CNLD status for each infant then allows us to test the discriminatory capacity of a probability value, for example ≥0.5, in predicting CNLD. For each individual value of probability selected, the test’s sensitivity and specificity can be calculated. It is then possible to plot the sensitivity and specificity for a range of probability values. A plot of sensitivity against 1-specificity is called a receiver operator characteristic (ROC) curve (fig 1).12-14 A ROC curve provides an alternative to single sensitivity and specificity values, allowing examination of a test’s discriminatory capacity across a full range of cut off values.

A predictive test with no false positive and no false negative results (a perfect test) would be represented by a line joining the points 0,0 and 1,1 (fig 1). A non-discriminatory test would give an equal number of false positive and negative results and would be represented by the diagonal line from 0,0 to 1,1. For example, tossing a two sided coin to predict CNLD would produce such a line.

Individual ROC curves can be given a value and compared by measuring the area under the curve. A perfect test would have an area of 1, a non-discriminatory test an area of 0.5.

The area under the ROC curve was used to represent the discriminatory capacity of the test over its full range of values. The standard error for this area was also calculated and the significance of any difference in the area under the curve was estimated.13

Given the rather subjective interpretation of mild pulmonary interstitial emphysema and terminal airway distension it was decided to

---

Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Leeds (n = 166)</th>
<th>Liverpool (n = 133)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (g)</td>
<td>1100 (480–1490)</td>
<td>1072 (580–1500)</td>
</tr>
<tr>
<td>Gestation (weeks)</td>
<td>26 (23–27)</td>
<td>28 (24–40)</td>
</tr>
<tr>
<td>Male sex</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>PDA</td>
<td>27</td>
<td>16</td>
</tr>
<tr>
<td>CNLD</td>
<td>28</td>
<td>44</td>
</tr>
<tr>
<td>Time ventilated (days)</td>
<td>5 (1–64)</td>
<td>8 (1–219)</td>
</tr>
</tbody>
</table>

PDA = patent ductus arteriosus.
also develop a model without the radiographic variable. This would also simplify the model and the final test.

**Results**

CNLD occurred in 47 (28%) of the Leeds neonates and in 59 (44%) of the Liverpool neonates. Table 1 gives the basic descriptive and clinical parameters. After the removal of variables which were not significantly independently associated with outcome, the following factors remained as explanatory variables: gestation, patent ductus arteriosus, ventilation status on day 7, maximum inspired oxygen fraction (FiO₂) between days 2 to 7, and abnormal gas pattern on radiographs.

The logistic regression equation for this model was

\[ y = 9 - 522 \times \text{gestational age} \\
+ 0.518 \times \text{maximum FiO₂,} \\
+ 2.603 \times \text{if patent ductus arteriosus} \\
+ 2.327 \times \text{if ventilated on day 7} \\
- 2.420 \times \text{barotrauma on radiograph} \]

This equation can produce a probability of CNLD for each infant by inserting y into the equation

\[ p = \exp(y)/(1 + \exp(y)) \]

For example, in a neonate of 28 weeks' gestation with a maximum FiO₂ of 80%, without a patent ductus arteriosus, still ventilated on day 7 but with no evidence of barotrauma, \( y = 1.745 \) and \( p = 0.85 \).

The incidence of pneumothorax and pulmonary interstitial emphysema was 12% and terminal airway distension was 40%. The logistic regression equation for this model without the radiographic variable was

\[ y = 7 - 624 \times \text{gestational age} \\
+ 0.423 \times \text{maximum FiO₂,} \\
+ 0.030 \times \text{if patent ductus arteriosus} \\
+ 1.107 \times \text{if ventilated on day 7} \\
+ 2.264 \times \text{if ventilated on day 7} \]

ROC curves were plotted for the data from the Leeds neonates using the two models (fig 1). Despite the removal of the radiological variable from the model, there was no significant reduction in the area under the curve (table 2). A similar curve was then constructed from data obtained for the Liverpool neonates, which showed a similar area under the curve, which was not significantly different (fig 2).

**Discussion**

The particular strengths of this predictive test were its requirement for easily recorded clinical variables and its validation in a different group of subjects from that in which it was derived. The reproducibility of the test, developed in Leeds neonates, was assessed in a population different from that from which it was derived, Liverpool neonates, and found to be equally effective. This occurred despite the higher incidence of CNLD in the Liverpool neonates. This performance compares favourably with earlier predictive tests, which really describe associations rather than specific attributes of the prediction.15 The ROC curves for the predictive test showed excellent discriminatory capacity. Their performance was significantly better than the C reactive protein concentration in predicting septicaemia in neonates.16

It is hoped to use the predictive test to identify neonates at high risk of CNLD to allow systematic intervention earlier, in an endeavour to improve outcome. The use of different levels of specificity allows the risk of any intervention to be limited appropriately. For instance, an intervention such as corticosteroid treatment is known to have significant risks. In this situation a specificity level of 97% may be chosen, although this would identify only 52% of neonates developing CNLD. In contrast with a less risky intervention, a probability value could be chosen with a high sensitivity (97%), but a low specificity (74%) (table 3).

The higher incidence of CNLD in Liverpool (44%) than Leeds (28%) means that, although the sensitivity and specificity of the predictive test were similar in the two groups, there were large differences between their positive and negative predictive values. For example, if a cut off probability value of 0-7 is chosen,

![Figure 2 ROC curves for the predictive test without the radiographic variable (XR) in Leeds neonates (in whom it was developed) and Liverpool neonates (in whom it was validated).](http://fn.bmj.com/ Arch Dis Child Fetal Neonatal Ed: first published as 10.1136/fn.71.1.F36 on 1 July 1994. Downloaded from http://fn.bmj.com/ on June 13, 2022 by guest. Protected by copyright.)

**Table 2 Area under the curve for the ROC curve for the two Leeds models and for the Liverpool neonates**

<table>
<thead>
<tr>
<th>Area under the curve</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leeds</td>
<td>0.913</td>
</tr>
<tr>
<td>Leeds, including radiograph</td>
<td>0.926</td>
</tr>
<tr>
<td>Liverpool</td>
<td>0.937</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>p Value</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥0.90</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>≥0.80</td>
<td>31</td>
<td>98</td>
</tr>
<tr>
<td>&gt;0.70</td>
<td>92</td>
<td>97</td>
</tr>
<tr>
<td>≥0.60</td>
<td>67</td>
<td>95</td>
</tr>
<tr>
<td>&gt;0.50</td>
<td>73</td>
<td>93</td>
</tr>
<tr>
<td>&gt;0.40</td>
<td>83</td>
<td>90</td>
</tr>
<tr>
<td>≥0.30</td>
<td>88</td>
<td>84</td>
</tr>
<tr>
<td>≥0.20</td>
<td>92</td>
<td>79</td>
</tr>
<tr>
<td>≥0.10</td>
<td>100</td>
<td>43</td>
</tr>
</tbody>
</table>

For example, this implies that, in Liverpool, 92% of the neonates would be expected to develop CNLD if their probability of developing it was ≤ 0.7, while 92% of those with a probability of development > 0.7 would not be expected to develop it.
sensitivity and specificity in the Leeds group were 55 and 96% respectively, compared with 52 and 97% in the Liverpool group. The positive predictive value, however, was 95% in the Liverpool group and 84% in Leeds neonates. The negative predictive value was 61% in the Liverpool neonates compared with 84% in the Leeds neonates.

The independent negative association between abnormal gas pattern and CNLD was surprising, given the positive historical association between them. The association was only significant if terminal airway distension was included. We had regarded the appearance of terminal airway distension, seen in 40% of neonates, as a sign of barotrauma, representing ventilator induced overinflation. We now believe that this appearance represents improving lung aeration and that we have seen the converse of a positive relation between CNLD and lung fluid retention, which is a well recognised association.

Consequently, abnormal gas pattern was removed from the model, with no significant reduction in the performance of the predictive tests, judged by the area under the ROC curve. By removing this variable, the model was simplified and made more objective.

One possible weakness of this predictive test was the use of variables which may have a subjective element to their value. Inspired oxygen fraction and ventilation status on day 7 are determined by both the underlying disease process and medical staff’s reaction to it. In the two neonatal units ventilatory parameters were adjusted in a standardised way, so that the subjective element was reduced to a minimum. It would be possible to improve the degree of objectivity by using parameters such as alveolar-arterial oxygen tension ratio or oxygenation index.

This predictive test should not be used to identify neonates above 1500 g birth weight who develop CNLD, as we only studied neonates below this birth weight. Given the small number of explanatory variables identified, we now plan to further validate the test for all neonates requiring ventilation.

Since the predictive test was developed there has been an increase in the use of antenatal steroids and surfactants. Despite this chronic lung disease remains a major problem. The test also needs to be validated in neonates who have received these treatments, as they are known to significantly alter some of the explanatory variables which we have chosen.

The outcome variable of this study was dependency on additional inspired oxygen at 28 days with an abnormal chest radiograph. There has been some debate about the relevance of this outcome in neonates born at, say, 24 weeks’ gestation, even though it is the predominant outcome measure in interventional studies. There have been suggestions that other outcomes are more important, including supplemental oxygen dependency at 36 weeks’ postconception and at discharge home. We aim to validate this technique against these outcomes in further developments of this predictive test.

Although the development of the prediction required sophisticated statistical analysis and assessment using ROC curves, the individual p values can be calculated using these formulas and a scientific calculator or a Basic computer language module, available on most personal computers. The advantage of this system is that it calculates an individual p value and so allows the test to be used at different levels of sensitivity and specificity.

We acknowledge the help of paediatric consultants in the north of England who provided follow up information.