CHAPTER 6

INDEPENDENT VALIDATION OF AN EXISTING MODEL ENABLES PREDICTION OF HEARING LOSS AFTER CHILDHOOD BACTERIAL MENINGITIS

Rogier C.J. de Jonge*
Marieke S. Sanders*
Caroline B. Terwee
Martijn W. Heymans
Reinoud J.B.J. Gemke
Irene Koomen
Lodewijk Spanjaard
A. Marceline van Furth

* Authors contributed equally

Submitted
ABSTRACT

Objective: This study aimed external validation of a formerly developed prediction model identifying children at risk for hearing loss after bacterial meningitis (BM). Independent risk factors included in the model are: duration of symptoms prior to admission, petechiae, cerebral spinal fluid glucose level, *Streptococcus pneumoniae* and ataxia. Validation helps to evaluate whether the model has potential in clinical practice.

Study design: A total of 116 Dutch school-age BM survivors were included in the validation cohort and screened for sensorineural hearing loss (>25 dB). Risk factors were obtained from medical records. The model was applied to the validation cohort and its performance was compared with the development cohort. Validation was performed by application of the model on the validation cohort and by assessment of discrimination and goodness of fit. Calibration was evaluated by testing deviations in intercept and slope. Multiple imputation techniques were used to deal with missing values.

Results: Risk factors were distributed equally between both cohorts. Discriminative ability (Area Under the Curve, AUC) of the model was 0.84 in the development and 0.78 in the validation cohort. Hosmer-Lemeshow test for goodness of fit was not significant in the validation cohort, implying good fit concerning the similarity of expected and observed cases. Sensitivity and negative predicted value were high, while specificity and positive predicted value were low which is comparable with findings in the development cohort.

Conclusions: Performance of the model remained good in the validation cohort. This prediction model might be used as an additional screening tool and can help to identify those children that need special attention and a long follow-up period or more frequent auditory testing.
INTRODUCTION

Due to successful vaccination programs, and to spontaneous decline in Neisseria meningitidis serogroup B infections, the incidence of bacterial meningitis (BM) in childhood is decreasing in the Western world. Still, BM ranks among the top ten causes of death in children younger than 14 years in high income countries. Further, developing countries account for 98% of the estimated 5.6 million disability-adjusted life years attributed to meningitis globally\(^1,2\). Sensorineural hearing loss is the most common severe consequence of BM, with an incidence in children of 7-31\%\(^3,4,5,6,7,8,9\). Hearing loss after BM is probably multicausal. Bacterial labyrinthitis due to dissemination of the infection from the subarachnoid space in combination with toxic or serous labyrinthitis, direct nerve fiber damage and secondary ischemic damage are thought to be part of the mechanism\(^10\).

Especially in (young) children even mild impairment in hearing abilities may impair auditory, linguistic, communication and learning skills. Early identification of hearing loss is indispensable for effective treatment resulting in the acquisition of normal linguistic development\(^11,12\). Further, cochlear ossification may complicate cochlear implantation making early diagnosis even more essential\(^13,14\).

The actual incidence of post-meningitis hearing loss is probably underestimated because audiometric testing is only performed in clinical suspected cases. Many cases are late or never diagnosed\(^9\). For that reason, routine hearing evaluation is recommended in the standard follow-up program of childhood BM aiming to achieve more timely intervention\(^8,14\).

To support the recognition of patients at high risk for hearing loss after BM, Koomen et al. developed a clinical prediction rule\(^6\). Clinical prediction rules are (regression) models that use three or more variables from patient history, clinical course or diagnostic tests to calculate a probability of an outcome measure. These rules are potentially strong tools that are currently used in clinical decision-making\(^15\).

The construction of the prediction rule for hearing loss was described in the original publication by Koomen et al.\(^6\). In short, 628 Dutch school-age children with a history of BM were included and screened for risk factors and for perceptive hearing loss. A prediction model was developed using multivariable logistic regression. Five risk factors for hearing loss were found to be independent determinants of hearing loss including: duration of symptoms prior to admission longer than two days, the absence of petechiae, CSF glucose level ≤0.6 mmol/L, \textit{S. pneumoniae} as causative pathogen and the presence of ataxia during
the illness. This model was transformed into a clinical prediction rule (figure 1). The scores and the matching probability of hearing loss were visually presented in a nomogram for use in clinical practice.

\[
\text{Risk score (max 100 points) } = 10 \times \text{symptoms > 2 days} + 27 \times \text{absence of petechiae} + 15 \times \text{CSF glucose level } \leq 0.6 \text{ mmol/L} + 12 \times \text{S. pneumoniae as causative pathogen} + 36 \times \text{ataxia as a symptom of BM}
\]

**FIGURE 1.** Clinical prediction rule for the prediction of hearing loss after childhood BM, as presented by Koomen et al.

In general a prediction model does not perform well in a different cohort than the one it was constructed in. For that reason external validation in an independent cohort is essential before implementation in practice. The aim of this study was to externally validate the existing model for hearing loss after childhood BM in a validation cohort of Dutch school-age BM survivors in order to evaluate the potential for usage in clinical practice.

**METHODS**

The study was approved by the medical ethics committee of the VU University Medical Center in Amsterdam. Written informed consent was obtained from the parents or guardians of all children and from the children themselves if they were at least 12 years of age.

**Development cohort**

The construction of the prediction rule was described in the original publication by Koomen et al. Files of the Netherlands Reference Laboratory for BM (NRLBM) were searched for children born between January 1986 and December 1994 who survived BM between January 1990 and December 1995. The NRLBM receives approximately 90% of the isolates of Dutch meningitis patients. Exclusion criteria included a complex onset of meningitis defined as: meningitis secondary to immunodeficiency states, CNS surgery, cranial trauma, CSF shunt infections of relapsing meningitis.

Sixteen hundred and five children were eligible for inclusion and their pediatricians were approached to send the parents a letter requesting participation. Six hundred and twenty-
Eight children were included, and their medical records were investigated for risk factors and for perceptive hearing loss of >25 dB. Forty-three children (7%) were diagnosed with hearing loss and a prediction model was developed using multivariable logistic regression and the aforementioned five predictors were selected. After internal validation using bootstrapping techniques and shrinkage of regression coefficients this model was transformed into a clinical prediction rule as presented in the figure 1. With this rule, a total score was calculated for each patient. The scores and the matching probability of hearing loss were visually presented in a nomogram for use in clinical practice.

Validation cohort
In 2005, the files of the NRLBM were searched again for Dutch children born between January 1993 and December 1999 who suffered from non-\textit{Haemophilus influenzae} type b (Hib) BM between January 1997 and December 2001. The exclusion criteria were identical to those used in the original study. One thousand and thirty-six children were eligible for inclusion, and the pediatricians were requested to send the parents (or guardians) an invitation letter to participate in the study. After informed consent the parents were sent screening questionnaires regarding health, learning and behavior. Three hundred and fifty eight children were included in the cohort used for validation of this model. Parallel to the approach used in the development study of Koomen \textit{et al.} in which a prediction rule for academic and behavioral problems after BM was constructed, a nested cohort approach was used. In this design only a subset of cases and controls are selected for further analysis, which decreases the necessary time and financial resources resulting in an improved efficiency.

In the total cohort of 358 children a nested cohort of 160 children were randomly selected and invited to visit our department for academic and neuropsychological testing and for the completion of questionnaires regarding behavior and health. Forty four of the invited children did not participate in this part of the study.

The outcome measure “hearing loss” was defined as a unilateral or bilateral perceptive loss of >25 dB and was based on findings in these records and on parental information provided in the questionnaires about the children’s health (Dutch versions of the Child Health Questionnaire (CHQ) and the Health Utilities Index (HUI) mark 2&3). Information on the risk factors for hearing loss was also collected by reviewing medical records of the pediatrician and the otolaryngologist after permission from the parents. Conductive hearing loss was not included. Finally, information about hearing loss was retrieved from 116 children. Figure 2 presents a flow chart of patient inclusion.
The risk factors evaluated for this study were the five aforementioned independent predictors included in the original model.

**Data analysis**

Univariate comparison of the distribution of patient characteristics and clinical data in the development and the validation cohort was performed by independent sample t-tests for continuous and $\chi^2$-tests for nominal data. Fisher’s Exact test was used if the data did not meet the criteria for a $\chi^2$-test. Statistical significance was considered with 2-tailed $p$-values of $\leq 0.05$.

**Missing data**

In both the development and the validation cohort multiple imputation techniques were used for substitution of missing variables. In the development cohort a maximum of 20%
of the data was missing per variable. In the validation cohort no more than 5% of the data was missing per variable, resulting in 9.5% missing cases for the whole model. Imputation was repeated five times, resulting in five different datasets.

**External validation**

For external validation the discriminative ability of the model in both cohorts was compared. In all sets the model with the original regression coefficients was applied on the data of each individual child. Receiver operating characteristic (ROC) curves of all imputed datasets were constructed for these results predicting the outcome measure “hearing loss”. An average area under the curve (AUC) was calculated and compared with the average AUC of the ROC curves of the development cohort. The fit of the model in the validation cohort was tested with the Hosmer-Lemeshow test for goodness of fit. For this test the patients were grouped by decile of predicted probability and differences between expected and observed outcome in the ten groups were tested using $\chi^2$-tests. The five results of these $\chi^2$-tests of the imputated datasets were pooled.

The intercept in a logistic regression model results in equal averages of the predicted probabilities and the outcomes when the model is applied in the validation cohort (“calibration in the large”). The slope reflects that the regression coefficients are correctly estimated and that they yield the correct predictions in the validation cohort. We also assessed more specifically whether calibration was successful by separately testing deviations in the intercept and calibration slope when the model was applied in the validation cohort.

For each child in the validation cohort the individual risk score was calculated. The distribution of the number of subjects across categories according to these scores were compared between the development and the validation cohort. Positive and negative predictive value, sensitivity and specificity were calculated for different cut-off points and were compared for their clinical value with those composed with the original prediction model.

All data was analyzed using SPSS Statistics 18.0 (IBM Corporation, Somers, NY) and R (The R Project for Statistical Computing).

**RESULTS**

Table 1 presents the distribution of the patient characteristics, independent predictors and the outcome measure “hearing loss” in the development and the validation cohort. Differences in distribution of characteristics between the two cohorts were observed for the duration of symptoms before admission, consciousness, presentation of the child with
### Table 1. Patient characteristics of the development and validation cohort

| Characteristics                                      | Validation cohort n = 116 | Development cohort n = 628 | P-value $f$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No$^a$</td>
<td>No$^a$</td>
<td>Development vs validation cohort</td>
</tr>
<tr>
<td>Age at infection (years) $^b$</td>
<td>116</td>
<td>628</td>
<td>0.58</td>
</tr>
<tr>
<td>Male gender $^c$</td>
<td>116</td>
<td>628</td>
<td>0.13</td>
</tr>
<tr>
<td>Duration of symptoms prior to admission (days) $^b$</td>
<td>113</td>
<td>621</td>
<td>0.001</td>
</tr>
<tr>
<td>Duration of symptoms &gt; 2 days $^c$</td>
<td>113</td>
<td>628</td>
<td>0.059</td>
</tr>
<tr>
<td>Seizures prior to admission $^c$</td>
<td>116</td>
<td>628</td>
<td>0.45</td>
</tr>
<tr>
<td>Seizures at admission $^c$</td>
<td>116</td>
<td>628</td>
<td>0.035</td>
</tr>
<tr>
<td>Decreased consciousness $^c$</td>
<td>106</td>
<td>601</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rectal temperature (°C) $^b$</td>
<td>100</td>
<td>583</td>
<td>0.64</td>
</tr>
<tr>
<td>Rectal temperature ≥ 38 °C $^c$</td>
<td>112</td>
<td>583</td>
<td>0.55</td>
</tr>
<tr>
<td>Meningeal irritation $^c$</td>
<td>109</td>
<td>628</td>
<td>0.018</td>
</tr>
<tr>
<td>Petechiae $^c$</td>
<td>115</td>
<td>618</td>
<td>0.67</td>
</tr>
<tr>
<td>Focal neurological deficits $^c$ $^d$</td>
<td>115</td>
<td>628</td>
<td>0.16</td>
</tr>
<tr>
<td>Middle ear infection $^c$</td>
<td>93</td>
<td>514</td>
<td>0.17</td>
</tr>
<tr>
<td>Laboratory tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSF leukocyte count (/μL)$^b$</td>
<td>111</td>
<td>548</td>
<td>0.90</td>
</tr>
<tr>
<td>CSF glucose (mmol/l) $^b$</td>
<td>110</td>
<td>543</td>
<td>0.94</td>
</tr>
<tr>
<td>CSF glucose ≤ 0.6 mmol/l $^c$</td>
<td>110</td>
<td>543</td>
<td>0.94</td>
</tr>
<tr>
<td>CSF protein level (g/l) $^b$</td>
<td>104</td>
<td>508</td>
<td>0.89</td>
</tr>
<tr>
<td>Causative pathogen in CSF: (total)</td>
<td>116</td>
<td>628</td>
<td></td>
</tr>
<tr>
<td>Pathogen</td>
<td>No. (%)</td>
<td>Control No. (%)</td>
<td>p-value</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>N. meningitidis</td>
<td>92 (79.3%)</td>
<td>495 (78.8%)</td>
<td>0.91</td>
</tr>
<tr>
<td>S. pneumonia</td>
<td>22 (19.0%)</td>
<td>103 (16.4%)</td>
<td>0.50</td>
</tr>
<tr>
<td>S. agalactie (group B)</td>
<td>2 (1.7%)</td>
<td>18 (2.9%)</td>
<td>0.76</td>
</tr>
<tr>
<td>E. coli</td>
<td>0 (0%)</td>
<td>10 (2.9%)</td>
<td>0.38</td>
</tr>
<tr>
<td>L. monocytogenes</td>
<td>0 (0%)</td>
<td>2 (0.3%)</td>
<td>1.0</td>
</tr>
<tr>
<td>Bloodculture positive for BM causing pathogen</td>
<td>116 49 (42.2%)</td>
<td>628 300 (47.8%)</td>
<td>0.27</td>
</tr>
<tr>
<td>Serum leukocyte count (x10^9) ^b</td>
<td>116 16.2 (9.4)</td>
<td>613 17.8 (10.7)</td>
<td>0.15</td>
</tr>
</tbody>
</table>

**Therapy and clinical course**

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. (%)</th>
<th>Control No. (%)</th>
<th>p-value</th>
<th>Abbrevations: No. = number, CSF = cerebrospinal fluid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dexamethasone prescribed ^c</td>
<td>110 19 (17.3%)</td>
<td>628 170 (27.1%)</td>
<td>0.029</td>
<td></td>
</tr>
<tr>
<td>Duration dexamethasone therapy (days) ^b</td>
<td>110 0.6 (1.5)</td>
<td>602 0.9 (1.8)</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>Mechanical ventilation ^c</td>
<td>116 14 (12.1%)</td>
<td>627 39 (6.2%)</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>Duration of hospitalization (days) ^b</td>
<td>115 12.9 (8.0)</td>
<td>618 14.7 (9.4)</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>Seizures during hospitalization ^c</td>
<td>115 8 (7.0%)</td>
<td>628 64 (10.2%)</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Duration of anti-epileptic therapy (days) ^b</td>
<td>112 0.6 (2.6)</td>
<td>610 1.4 (5.7)</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>Focal neurological deficits ^c,d</td>
<td>114 7 (6.1%)</td>
<td>627 94 (15.0%)</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td><strong>(transient) ataxia ^c,e</strong></td>
<td>115 3 (2.6%)</td>
<td>628 16 (2.5%)</td>
<td>0.97</td>
<td></td>
</tr>
</tbody>
</table>

**Outcome measure**

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. (%)</th>
<th>Control No. (%)</th>
<th>p-value</th>
<th>Abbrevations: No. = number, CSF = cerebrospinal fluid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hearing loss</td>
<td>116 14 (12.1%)</td>
<td>628 43 (6.8%)</td>
<td>0.052</td>
<td></td>
</tr>
</tbody>
</table>

^a Number of subjects the variable was obtained
^b Mean (standard deviation)
^c Number of subjects (%)
^d Focal neurological deficits are defined as cranial nerve deficits, increased or decreased reflexes of arms or legs, increased or decreased tonus of arms or legs, focal convulsions and ataxia.
^e (Transient) ataxia was defined as signs of ataxia, which lasted at least until discharge from the hospital, as documented in the medical records.
^f P-value: independent sample t-test for continuous data; χ^2^-test for nominal data, or Fisher’s Exact test if the data does not meet the criteria for χ^2^-test.

Abbrevations: No. = number, CSF = cerebrospinal fluid
meningeal irritation at admission, (duration of) dexamethasone prescription, number of children mechanically ventilated, duration of hospitalization, duration of anti-epileptic therapy and the number of children with focal neurological deficits. There were no significant differences in the distributions of the five predictors of the prediction rule. The incidence of hearing loss was 6.8% in the development cohort and 12.1% in the validation cohort.

**External validation of the prediction rule**
The average AUC of the ROC curves of the validation cohort was 0.78 (95% CI 0.64-0.92). In the development cohort, the reported AUC was 0.84 (95% CI 0.78-0.91) \(^6\). The pooled result of the Hosmer-Lemeshow test for goodness of fit of the five datasets was not significant (\(p\)-value 0.10), indicating good fit of the model in the validation cohort concerning the similarity of expected and observed cases of hearing loss. When the original model was applied in the validation cohort there was no significant deviation in intercept (\(p=0.07-0.11\)) and calibration slope (\(p=0.14-0.34\)).

**Comparison of the distribution of risk score categories**
Table 2 shows the numbers of subjects across categories of the score for both the development and the validation cohort. Children without hearing loss had an average risk score of 19.6 and children with hearing 45.3. In both cohorts approximately 40% of children without hearing loss score zero points. Only 1 child of all children with hearing loss has a score of 0. While most of the children with hearing loss had a higher score, children without hearing loss did have high scores as well. Table 3 presents positive and negative predictive value, sensitivity and specificity for the different cut-off points. Using a cut-off point of zero, the sensitivity is 100% while at a score of 1-25 the negative predicted value is 97.8%. At a maximum score on the risk score the specificity is 96.1%, but decreases rapidly at a lower score. The positive predictive value is low (60%). The conclusion of these tables is that sensitivity and negative predictive value are good, but specificity and positive predictive value are poor.
TABLE 2. Distribution of children with and without hearing loss across categories of the risk score

<table>
<thead>
<tr>
<th>Risk score</th>
<th>Development</th>
<th>Validation</th>
<th>Development</th>
<th>Validation</th>
<th>Development</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score 0</td>
<td>236 (38)</td>
<td>45 (39)</td>
<td>0 (0)</td>
<td>1 (7)</td>
<td>236 (40)</td>
<td>44 (43)</td>
</tr>
<tr>
<td>Score 1-24</td>
<td>86 (14)</td>
<td>322 (51)</td>
<td>14 (12)</td>
<td>59 (51)</td>
<td>2 (5)</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Score 25-36</td>
<td>87 (14)</td>
<td>409 (65)</td>
<td>15 (13)</td>
<td>74 (64)</td>
<td>5 (11)</td>
<td>7 (16)</td>
</tr>
<tr>
<td>Score 37-63</td>
<td>197 (31)</td>
<td>606 (96)</td>
<td>32 (27)</td>
<td>106 (91)</td>
<td>24 (56)</td>
<td>31 (72)</td>
</tr>
<tr>
<td>Score ≥64</td>
<td>22 (3)</td>
<td>628 (100)</td>
<td>10 (9)</td>
<td>116 (100)</td>
<td>12 (28)</td>
<td>43 (100)</td>
</tr>
</tbody>
</table>

TABLE 3. Positive and negative predictive value, sensitivity and specificity for the different cut-off points of the prediction rule

<table>
<thead>
<tr>
<th>Risk score cut-off point</th>
<th>Positive predictive value (%)</th>
<th>Negative predictive value (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score ≥0</td>
<td>6.8</td>
<td>12.1</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Score ≥1</td>
<td>11.0</td>
<td>18.3</td>
<td>100.0</td>
<td>92.9</td>
</tr>
<tr>
<td>Score ≥25</td>
<td>13.4</td>
<td>19.3</td>
<td>99.4</td>
<td>97.8</td>
</tr>
<tr>
<td>Score ≥37</td>
<td>16.4</td>
<td>23.8</td>
<td>98.3</td>
<td>94.9</td>
</tr>
<tr>
<td>Score ≥64</td>
<td>54.5</td>
<td>60.0</td>
<td>94.9</td>
<td>92.5</td>
</tr>
</tbody>
</table>

PREDICTION OF POST-MENINGITIS HEARING LOSS
DISCUSSION

In this study, the prediction model for hearing loss after childhood BM was validated successfully in a new independent cohort of school-age children. This is the first validated model for the identification of children at high risk for hearing loss after BM. We show that the risk score reaches excellent sensitivity and negative predictive value at a cut-off point of zero. Our recommendation is that at least the patients who are positive at one or more risk score should achieve timely and frequent hearing evaluation. It is not a replacement for standard hearing tests, but an addition to these tests, to estimate the risk for hearing loss in an early stage of disease.

Despite increasing awareness and recommendation of routine hearing evaluation in the standard follow-up of childhood BM, the amount of children whom hearing is not tested or severe hearing loss is even missed could be up to 25-30%. Relatively little is known about the clinical course of hearing loss after BM. It is not always noticeable or present directly after the infection and fluctuation and deterioration of hearing later in time might occur. Modern follow-up protocols therefore include a prolonged period of hearing loss evaluation. The only current treatment option in complete and profound hearing loss is cochlear implantation which may only be possible in a critical period. A good opportunity for hearing restoration could disappear within weeks since labyrinthitis ossificans makes implantation extremely difficult or even impossible. This prediction model can be used as a screening tool and can help to identify those children that need special attention and a long follow-up period or more frequent auditory testing even when the first test is negative.

For clinical practice the optimal cut-off point of the risk score has to be defined. Conform Koomen et al. we state that hearing loss should not be missed at all and therefore propose a low cut-off value of zero points. In this study it was confirmed that this is the optimal cut-off point. The risk score reaches excellent sensitivity and negative predictive value at this cut-off point. Unfortunately, as one increased the sensitivity of a rule, its specificity tends to decrease (and vice versa). In clinical practice, this indicates that our prediction rule has to be used to select the children at high risk, accepting the fact that a relatively high number of children without hearing impairment will be selected as well. Although with a low cut-off point sensitivity is high, in the validation cohort one child with a score of zero points did have hearing loss, while in the construction cohort no case was missed. We conclude that at least those patients who are positive at one or more factor score should achieve timely and frequent hearing evaluation. In the lower risk group, hearing must be evaluated as well, but it may be considered to perform auditory tests less frequently or for a shorter time period. This may improve the balance between costs and benefit. It can
create awareness of the importance of hearing evaluation in clinicians and the parents in an early stage of disease. The prediction rule must be easy to access, for instance on a website where clinicians can easily fill in the variables and the risk is calculated.

This study has several strengths. This is the first model developed and externally validated that predicts hearing loss after BM in childhood. In the last decade, many prediction models have been designed and presented in literature. Before implementation in practice is possible, external validation is an essential step\textsuperscript{16, 17, 18}. However, the majority of proposed models have never been validated\textsuperscript{18, 32}. Therefore, successful validation of our model is an important step forward in the development of a complete protocol for the follow-up of children who suffered from BM.

In general, the discrimination of a prediction model is interpreted to be fair, good or excellent when the AUC is or 0.7-0.8, 0.8-0.9 or 0.9-1.0, respectively. This model has an AUC of 0.84 in the development cohort and an AUC of 0.78 in the validation cohort, thus can be at least considered as fair.

With the help of the files from the NRLBM we could approach 90% of all Dutch pediatric meningitis patients, resulting in a highly representative patient group in the Netherlands. Further, state of the art methodology is used and described clearly. Therefore it may be easily to reproduce in a wide range of future studies.

An important possible limitation include the risk of selection bias. From the 1036 patients selected from the NRLBM database, 116 were included. Most children were excluded because approached pediatricians or parents refused participation or because the parents could not be contacted due to missing or incorrect address data, which, in our opinion this did not lead to selection of a specific group of patients. From the resulting 361 children 116 were included for validation of the hearing loss model. The nested cohort approach was used, while in the original design of the hearing loss rule the total cohort was used\textsuperscript{6, 20}. This design, in which only a subset of cases and controls are randomly selected for further analysis, is a known and appreciated methodology that results in more efficiency with reference to time and financial resources\textsuperscript{21}. This stepwise construction may increase the risk of selection bias leading to differences in case mix.

There is a difference in case mix and in the incidence of the outcome measure between the development and validation cohorts. The difference in the incidence of sensorineural hearing loss between the two cohorts was 6.8% in the development versus 12.1% in the validation cohort ($p$-value=0.052). This remarkable finding may be explained by the fact that in the development cohort 27% of the children were not tested for hearing loss, resulting in underestimation of the incidence\textsuperscript{6}. Further, an increase in awareness in
the period of approximately seven years between the two studies might have resulted in increased follow-up and incidence of hearing loss. Other case mix differences were caused by differences in variables, such as: duration of symptoms >2 days, meningeal irritation, dexamethasone therapy, mechanical ventilation, duration of anti-epileptic therapy and focal neurological deficits. These differences could have been responsible for a smaller spread in the risk for hearing loss (0-91% in the development versus 2-70% in the validation cohort). This could in turn be responsible for the small drop in discrimination of the model in the validation cohort. The performance of the model remained good, this may have not affected the model.

Information bias may have occurred. Risk factors and the outcome were determined retrospectively, with a risk of bias and missing data. To minimize the chance of bias, collection and interpretation of clinical data from the medical files was performed in a standardized way, and consensus meetings were held when there was any doubt about how to interpret the medical files. The diagnosis “perceptive hearing loss of >25 dB” is made in the Netherlands by clinicians or audiologists by standard tests. Multiple imputation techniques were used for substitution of missing data, which is currently the most reliable strategy to deal with missing information on covariates.

Further, sample size of the validation cohort was small. A frequently used general rule in the development of clinical prediction models is that the ratio events per variable (EVP) should be approximately 10. In the development cohort there were 43 events for 5 predictors included in the model, resulting in an EVP of 8.6. For external validation new insights in sample size issues are developed in recent years. It has been suggested that to detect minor changes in discrimination and calibration, and prevent type II error (the Null hypothesis of equal model performance is falsely not rejected), the validation cohort should contain at least 100 events and 100 non-events. The sample size of our validation cohort is limited, and therefore small differences in calibration might have been missed.

It is hypothesized that the distribution of BM pathogens has changed since the introduction of these vaccines and therefore the model will not perform well in the contemporary population. Since the development of the prediction rule two new vaccines against N. meningitidis serogroup C and high frequency serotypes of S. pneumoniae were introduced in the Netherlands and other western countries: Accompanied by a spontaneous decrease of N. meningitidis serogroup B infections this resulted in a dramatic decrease in incidence of BM. Serotype replacement with increasing incidence of infections has not yet occurred for N. meningitidis, but new outbreaks with non-vaccine-serotypes are possible and have occurred in history. For S. pneumoniae there is a net reduction of invasive infections but an increase in infections with non-vaccine-serotypes is observed. Therefore, we
should be careful expecting BM to become and stay a rare disease, and prediction rules
for sequelae remain valuable. In a recent study our group simulated a population without
the seven serotypes of *S. pneumoniae* included in the vaccine, and showed that the model
remains stable in this situation. To investigate whether the long term consequences of
vaccination have impact on the performance of the model, another validation study in
children that were infected more recently is planned.

Another important development is the rapidly increasing knowledge on the influence
of genetics on the course of diseases. This also applies for BM and hearing loss, and in
recent publications an effort was made to disclose the influence of genetic variation in
the immune response and the course of the disease in children. Van Well *et al.* described
a strong association between host genetic polymorphisms in pathogen recognition
receptors and hearing loss after BM extracted from both development and validation
cohort described in this study. Additional analyses will be performed to study whether
our model can be extended with genetic factors to predict post-meningitis hearing loss
even more accurately.

We conclude that we created and externally validated a clinical useful tool in addition
to regular auditory testing for the identification of children at high risk for hearing loss
after BM. In the future larger development and validation studies must be performed,
in which hearing loss is measured prospectively and genetic risk factors are included in
the construction of the prediction model. In our opinion, international cooperation is the
answer to the problem of decreasing incidence in the development of prediction models.
CHAPTER 6

REFERENCES


