Prospective Testing of Two Models Based on Clinical and Oximetric Variables for Prediction of Obstructive Sleep Apnea*

Nicolas Roche, MD; Bertrand Herer, MD; Catherine Roig, MD; and Gérard Huchon, MD, FCCP

Study objective: To test the validity of two models for prediction of obstructive sleep apnea syndrome (OSAS) before polysomnography.

Design: Prospective study.

Setting: Sleep laboratory in an obesity clinic.

Patients: Data from two populations were analyzed: the first (group 1) included 102 consecutive overweight patients referred to our laboratory by an obesity clinic between May 1992 and November 1994, and was used to develop the prediction models. The second (group 2) included 108 consecutive new patients referred to our laboratory by the same obesity clinic between February 1997 and September 1998, and was used to test the prediction models.

Measurements and results: Models were developed using a clinical score, pulmonary function tests, arterial blood gas tensions, and nocturnal pulse oximetry. OSAS was defined by an apnea-hypopnea index (AHI) > 15 events per hour, as measured by full-night polysomnography. Step-by-step multiple linear regression analysis (MLR) was used to provide an equation for calculation of predicted AHI, while logistic regression analysis (LR) provided an equation for calculation of the probability (P') of having OSAS. Characteristics of groups 1 and 2 were similar except for the prevalence of OSAS, which was higher in group 2 (74% vs 39% in group 1). The negative predictive value (NPV) of the MLR model dropped from 82.9% in group 1 to 36.7% in group 2. In parallel, the NPV of a P' < 0.25 according to LR decreased from 78.6% in group 1 to 23.5% in group 2.

Conclusion: Our results emphasize the need for systematic prospective testing of mathematical predictive models in OSAS, since their diagnostic characteristics may differ markedly between populations, even when the setting and mode of recruitment remain unchanged.

(CHEST 2002; 121:747–752)

Key words: case-finding; clinical score; nocturnal obstructive sleep apnea syndrome; oximetry; polysomnography; prediction models

Abbreviations: AHI = apnea-hypopnea index; BMI = body mass index; CS = clinical score; CT<sub>80</sub> = cumulative time spent below 80% of arterial oxygen saturation; LR = logistic regression analysis; MLR = multiple linear regression analysis; NPV = negative predictive value; OSAS = obstructive sleep apnea syndrome; PPV = positive predictive value

The “gold standard” for the diagnosis of obstructive sleep apnea syndrome (OSAS) is polysomnography, which is expensive and may not be easily accessible. However, the number of patients referred to sleep laboratories is likely to increase, due to the estimated prevalence of the disease (ie, 1 to 4% in middle-aged adults) and the increased knowledge of practitioners.

Therefore, several strategies have been developed to decrease the number of polysomnographic studies in patients with suspected OSAS; such strategies include ambulatory polysomnography, nap or half-night polysomnography, simplified multichannel recordings, snoring analysis, morphometric assessment, nocturnal oximetry, and multivariate prediction rules developed using mathematical modeling and, more recently, neural networks. How-
ever, only a few of these prediction rules have been tested prospectively in populations other than those in which they have been developed,3–6 which raises concern regarding their external validity. In addition, prospective testing has usually been performed immediately after the development of the prediction model, which limits the risk of changes in the studied population. Finally, some of these prediction models appear too complex for routine clinical use.

In a previous study, we used multivariate logistic regression analysis (LR) to develop a mathematical model for prediction of the probability of having OSAS.7 Nineteen variables were prospectively recorded, including clinical characteristics, arterial blood gas tensions, and pulmonary function and nocturnal oximetry variables. The analysis yielded two equations: the first equation was based on all variables, and the second equation was based only on the three independent predictors of OSAS, ie, gender, cumulative time spent below 80% of arterial saturation (CT So2), and a clinical score (CS) based on snoring, reported interrupted nocturnal breathing, excessive daytime sleepiness, and arterial hypertension. The negative predictive value of a calculated probability (P') of having OSAS of < 0.25 was 94% with the first equation, and 78.6% with the second equation.7

To determine the external validity of such prediction models, we prospectively studied 108 additional consecutive new patients referred to the same sleep center. Since the first LR equation was believed to be far too complex to be used in routine clinical practice, we tested only the second equation. In addition, since the overall accuracy of this logistic regression model in group 1 was not very high (66%), we tested whether a linear regression model for prediction of the apnea-hypopnea index (AHI) could provide more accurate and reproducible results. Therefore, we used the same 19 variables from the first population to develop a multiple linear regression (MLR) equation for calculation of predicted AHI, and tested the diagnostic characteristics of this model in the second group of patients.

### Materials and Methods

#### Study Populations

We analyzed data from two populations in which a CS was established and pulmonary function tests, nocturnal oximetry, and full-night polysomnography were performed. The first population (group 1) included 102 consecutive overweight patients referred to our laboratory by an obesity clinic between February 1997 and September 1998, and was used to test the prediction models.7

The second population (group 2) included 108 consecutive new patients referred to our laboratory by the same obesity clinic between February 1997 and September 1998, and was used to test the prediction models.

#### Recorded Variables

Details of measurements and analysis in group 1 have been reported previously.7 In group 2, recorded variables were restricted to clinical score, nocturnal oximetry and polysomnography. The CS was the sum of four variables (each of which was scored 0 or 1): habitual snoring, interrupted nocturnal breathing as reported by the spouse or roommates, excessive daytime sleepiness, and arterial hypertension. Body mass index (BMI) was not included since it was ≥ 25 kg/m² in all patients of group 1.7 Overnight pulse oximetry was performed 1 to 7 days before full-night polysomnography; which was performed and analyzed as previously described in all patients (SEFAM Minisomn; Mallinckrodt; Nancy, France, or Medatec; Brussels, Belgium).7 OSAS was defined as an obstructive AHI of at least 15 events per hour.5,9

#### Statistical Analysis

The method used to develop the LR prediction model has been previously described.7 Briefly, logistic regression analysis was used to calculate the probability of having OSAS using the following equation: 

\[ P = e^y/(1 + e^y) \]

where \( y = c + x_1 \text{variable 1} + x_2 \text{variable 2} + \ldots \) The constant (c) and parameter estimates (\( x_1, x_2, x_3, \ldots \)) were determined by a multivariate analysis in which presence or absence of OSAS was the dependent variable, and in which the step-by-step Wald method was used to restrain variables to those that independently correlated to the presence of OSAS. A second prediction model was developed using MLR, AHI being the dependent variable and all other variables from group 1 being introduced in the equation in a stepwise manner. Simple linear regression analysis was used to study correlations between measured and calculated AHI, and between measured AHI and calculated probability. A visual analysis of corresponding plots allowed to identify cutoffs that could be useful to predict the presence or absence of OSAS. We then used receiver operating characteristic curves to calculate the diagnostic characteristics (ie, sensitivity, specificity, negative predictive value [NPV], positive predictive value [PPV], and overall accuracy) of the two equations in groups 1 and 2. NPV is the probability of being free of the disease if the test result is negative (true-negatives/true-negatives plus false-negatives), while PPV is the probability of having the disease if the test result is positive (true-positives/true-positives plus false-positives); overall accuracy is the proportion of cases adequately classified as having the disease or not (true-positives plus true-negatives/all subjects). Comparisons between groups were performed using Student’s t test. Results are expressed as mean ± SEM unless indicated. We considered p < 0.05 to be significant. Statistics were performed with software (BMDP Statistical Software; Los Angeles, CA) and SPSS (SPSS; Chicago, IL); receiver operating characteristic curve analysis was performed with ROC Analyzer software (R. M. Centor and J. Keightley; University of Alabama, Birmingham, AL).

#### Results

#### Characteristics of the Populations

Age, gender, and BMI were similar in both groups (Table 1). All patients with an AHI > 15/h had predominantly obstructive apneas and were therefore classified as having OSAS, which was signifi-
cantly more frequent in group 2 than in group 1. Although the severity of OSAS, as assessed by AHI, was similar in the two groups, the CS and CT80 of patients with OSAS were higher in group 1 than in group 2.

Prediction Models Developed in Group 1

The following equation was used for calculation of the probability of having OSAS: 

\[ P = \frac{e^y}{1 + e^y}, \]

where

\[ y = -4.31 + (1.41 \times \text{sex}) + (0.03 \times \text{CT}_{80}) + (0.78 \times \text{CS}). \]

Details on the development of this model and adjusted odds ratio for each independent variable have been previously reported.7 The equation for calculation of predicted AHI, as obtained by MLR in group 1, was as follows:

\[ \text{AHI} = -7.24 + (7.04 \times \text{CS}) + (0.38 \times \text{CT}_{80}). \]

Correlation coefficients with AHI were 0.38 for CS (p = 0.0002) and 0.41 for CT80 (p < 0.0001).

Diagnostic Characteristics of the Models in Groups 1 and 2

The probability of having OSAS, as calculated by LR, correlated to AHI only in group 1, as shown in Figure 1. The diagnostic characteristics of the LR prediction model were different in groups 1 and 2, with a lower overall accuracy in group 2 (53% vs 66% in group 1); more specifically, NPV was much lower and PPV was higher in group 2 than in group 1 (Fig 2).

As shown in Figure 3, measured AHI and AHI calculated by MLR were significantly correlated in groups 1 and 2. However, dispersion was greater in group 2, as reflected by a lower correlation coefficient (r = 0.513 in group 1 vs r = 0.381 in group 2). The overall accuracy of the MLR model was similar in both groups (65.7% in group 1 and 62.1% in group 2). However, diagnostic characteristics of this model were different in group 2 when compared to group 1.

### Table 1—Characteristics of the Populations

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Group 1 (n = 102)</th>
<th>Group 2 (n = 108)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yr</td>
<td>53.20 ± 1.08</td>
<td>53.30 ± 1.05</td>
</tr>
<tr>
<td>Proportion of OSAS, %</td>
<td>39</td>
<td>74*</td>
</tr>
<tr>
<td>Gender, % men</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>OSAS</td>
<td>85</td>
<td>72.5</td>
</tr>
<tr>
<td>Non-OSAS</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>39.57 ± 0.93</td>
<td>39.50 ± 0.9</td>
</tr>
<tr>
<td>OSAS</td>
<td>39.68 ± 1.10</td>
<td>39.70 ± 1.03</td>
</tr>
<tr>
<td>Non-OSAS</td>
<td>39.5 ± 1.66</td>
<td>38.88 ± 1.70</td>
</tr>
<tr>
<td>Clinical score</td>
<td>2.56 ± 0.10</td>
<td>2.90 ± 0.08</td>
</tr>
<tr>
<td>OSAS</td>
<td>3.28 ± 0.13</td>
<td>2.94 ± 0.10*</td>
</tr>
<tr>
<td>Non-OSAS</td>
<td>2.60 ± 0.13</td>
<td>2.79 ± 0.16</td>
</tr>
<tr>
<td>CT80, %</td>
<td>14.57 ± 2.16</td>
<td>8.07 ± 1.67*</td>
</tr>
<tr>
<td>OSAS</td>
<td>23.27 ± 4.05</td>
<td>9.46 ± 2.00*</td>
</tr>
<tr>
<td>Non-OSAS</td>
<td>9.44 ± 2.18</td>
<td>4.1 ± 2.79</td>
</tr>
<tr>
<td>AHI, per h</td>
<td>19.26 ± 2.23</td>
<td>33.22 ± 2.37*</td>
</tr>
<tr>
<td>OSAS</td>
<td>42.7 ± 2.98</td>
<td>41.9 ± 2.57</td>
</tr>
<tr>
<td>Non-OSAS</td>
<td>4.13 ± 0.50</td>
<td>8.5 ± 0.62</td>
</tr>
</tbody>
</table>

*OSAS = obstructive sleep apnea syndrome; BMI = body mass index; CT80 = cumulative time spent below 80% arterial oxygen saturation; AHI = apnea + hypopnea index.

*p < 0.05 for comparison between group 1 and group 2.
Discussion

Our data demonstrate that two regression models for prediction of OSAS developed in a first group of patients exhibit markedly different diagnostic characteristics in a second group recruited a few years later. It is unlikely that these differences relate to marked modifications in criteria for patients selection, since the mode of recruitment and baseline characteristics were similar in the two groups.

The aim of mathematical models for prediction of OSAS is to decrease the number of polysomnographic studies by providing high NPV and/or PPV in some subsets of patients. These two variables usually evolve in opposite directions, as observed in our two groups of subjects: in the first group, the two models had relatively high NPV but quite low PPV, while the opposite was found in the second group. Such differences in diagnostic characteristics raise doubts about the utility of such models in clinical practice, since a high proportion (68% with MLR and 76% with LR) of patients in whom polysomnography would not be performed in the second group would be misdiagnosed as not having OSAS.

NPV and PPV depend not only on the sensitivity and specificity of the test, but also on the prevalence of the disease in the studied population, which was much higher in the second group than in the first group. This discrepancy may in part explain the variation in diagnostic characteristics of the models. Since the mode of recruitment was exactly the same for the two groups of subjects, the most likely explanation for the variation in the prevalence of OSAS is an improvement in the knowledge of clinical presentations of OSAS by physicians during the delay between the recruitment of the two groups (ie, 3 years), leading to a better selection before referral for polysomnography.

The variations in diagnostic characteristics between the two groups may also relate to a too-low diagnostic performance of the mathematical models (since the overall accuracy of both models was only 66% in the first population), which could in part be due to a too-low number of subjects in group 1. Similar accuracies have been found in most studies based on clinical variables, combined or not with oximetry data.5,10–16 However, recent models devel-
developed by neural networks appear to have a higher accuracy.\textsuperscript{3,4} Such models are interesting since they can maintain or increase their performance by regularly recalculating predictive equation according to newly integrated results. In our study, the relatively low accuracy was likely due to the limited number of variables (n = 6) introduced in these models (including the four variables of the CS, gender, and CT\textsubscript{50}). Indeed, in our previous study, the 19-variable logistic regression model was more accurate than the 3-variable model, since both its NPV and PPV values were \( \geq 90\% \) (NPV of P' < 0.25, 94%; PPV of P' > 0.75, 90%).\textsuperscript{7} However, we believe that such a high number of variables would make this model much too complex to be used in routine clinical practice, especially because some of the variables (ie, indexes of upper-airway obstruction) are not automatically calculated by most spirometers. Accordingly, we decided to test only the simplified equation. A probability of having OSAS of < 0.25, as calculated by this equation, had a NPV of 78.6\% in the first population; using such a model in this population would have allowed to avoid 13\% of polysomnography without “missing” > 10\% of patients with OSAS. Such a performance is obviously suboptimal but may have been considered as acceptable (ie, better than no selection at all) in settings where the availability of polysomnography is limited, provided that its NPV remained stable over time, which does not appear to be the case.

We did not include the \( \Delta \) index (a saturation variability index) or the number of \( \geq 4\% \) dips in arterial oxygen saturation in the studied criteria, since these variables had not been extensively studied when the first study was designed. Such indexes may increase the diagnostic yield of oximetry\textsuperscript{8,9,17}; however, they remain to be validated since some studies found that oximetry is not a cost-effective tool for case-finding in OSAS.\textsuperscript{18} We also did not include neck circumference in our models. It has been advocated that this variable is more closely associated with OSAS than BMI, although data on this topic are controversial.\textsuperscript{6,19,20} Our patients were all overweight (ie, BMI > 25 kg/m\(^2\)), so that BMI did not correlate to AHI and did not differ between subjects with and without OSAS; therefore, it was not included in the prediction models. We cannot rule out that neck circumference could have been more discriminant.

In conclusion, our results emphasize that, even in a stable setting with unchanged mode of recruitment, the diagnostic characteristics of OSAS prediction models based on clinical and oximetric variables

\begin{table}
\centering
\caption{Diagnostic Characteristics of Multiple Linear Regression Equation\textsuperscript{*}}
\begin{tabular}{|l|c|c|}
\hline
Characteristics & Group 1 & Group 2 \\
\hline
Sensitivity & 82.5 & 61.3 \\
Specificity & 55.0 & 64.3 \\
NPV & 82.9 & 56.7 \\
PPV & 54.1 & 83.1 \\
Overall accuracy & 65.7 & 62.1 \\
\hline
\end{tabular}
\textsuperscript{*}Data are presented as %. \\
PPV = positive predictive value; NPV = negative predictive value.
\end{table}

\begin{figure}[h]
\centering
\begin{tabular}{|c|c|}
\hline
\textbf{group 1} & \textbf{group 2} \\
\hline
\includegraphics[width=\textwidth]{group1} & \includegraphics[width=\textwidth]{group2} \\
\hline
\end{tabular}
\caption{Relationship between measured AHI and AHI calculated using MLR. \textit{Top:} group 1, in which the prediction model was developed (\( r^2 = 0.263, p < 0.0001 \)). \textit{Bottom:} group 2, in which the model was tested (\( r^2 = 0.146, p < 0.0001 \)).}
\end{figure}
may vary markedly over time; this raises concerns about the predictive usefulness of these variables and the external validity of such models, which should be systematically retested in other populations.

REFERENCES
1. Systematic review of the literature regarding the diagnosis of sleep apnea. Evid Rep Technol Assess 1999; (Summer):i-154