Prediction of Nonrecovery in Bell’s Palsy Using Sunnybrook Grading

Elin Marsk, MD; Nina Bylund, MD; Lars Jonsson, MD, PhD; Lalle Hammarstedt, MD, PhD; Mats Engström, MD, PhD; Nermin Hadziosmanovic, MSc; Thomas Berg, MD, PhD; Malou Hultcrantz, MD, PhD

Objectives/Hypothesis: To develop a clinical prognostic model to identify Bell’s palsy patients with risk for nonrecovery at 12 months.

Study Design: Data from a prospective, randomized, double-blind, placebo-controlled, multicenter study.

Methods: There were 829 patients with Bell’s palsy randomized in a factorial fashion to treatment with prednisolone or no prednisolone. Facial function was assessed with the Sunnybrook grading scale. Univariate and multivariate logistic regression analyses at different time points were used to identify factors predicting nonrecovery, defined as Sunnybrook score <70 at 12 months. Variables studied were age, gender, time to inclusion, prednisolone treatment, side of palsy, pain at inclusion, and Sunnybrook scores. Factors of predictable significance were used to construct prognostic models at baseline, days 11 to 17, and at 1 month. Receiver operating characteristics curves were created to test the predictive capacity of the models.

Results: At baseline, treatment with prednisolone or no prednisolone (P = .0005), age (P = .04) and the Sunnybrook score (P = .0002) were significant factors for predicting nonrecovery. The receiver operating characteristics area under the curve at baseline for these three variables was 0.74 (sensitivity 0.83, specificity 0.57). At days 11 to 17 and at 1 month, the Sunnybrook score was the only significant predictive variable. The respective areas under the curves for the Sunnybrook score at these time points were 0.83 (sensitivity 0.81, specificity 0.75) and 0.94 (sensitivity 0.91, specificity 0.85).

Conclusions: Sunnybrook grading at 1 month most accurately predicts nonrecovery at 12 months in Bell’s palsy.

Key Words: Facial palsy, facial paralysis, facial nerve, prognosis, multivariate analysis, receiver operating characteristics, area under the curve.

Level of Evidence: 1b

INTRODUCTION

Bell’s palsy is an acute unilateral peripheral facial palsy or paralysis in which no etiology, such as infection, neoplasm, autoimmune disease, or trauma can be identified. The disease is the most common form of acquired peripheral palsy, with approximately 30 patients per 100,000 individuals and year affected. Its highest incidence occurs between 15 to 45 years of age. Bell’s palsy varies in clinical severity and recovery. Prednisolone treatment improves recovery rates, but even with this treatment, 6% to 27% of patients will suffer sequelae with facial asymmetry, contracture, and synkinesis of the mimic muscles.

To identify patients with a poor recovery prognosis and those who may benefit from physiotherapy, possible future medical treatments other than prednisolone, or early surgical intervention, new models for clinical prognostication are needed. Such models can also increase the power of therapeutic studies by adjusting for prognostic factors and also individualize therapy in accordance with the expected outcome.

Electroneurography is the most powerful tool for prognostication of nonrecovery in Bell’s palsy. In the clinical setting, however, access to neurophysiological tests are often limited. As a result, factors that can predict prognosis have been sought in studies based on clinical evaluation of acute symptoms and signs, concurrent medical diseases, age, and radiological investigations. However, these predictive factors are not powerful enough in the clinical situation, and additional clinical prognostic tools are therefore needed.

The Scandinavian Bell’s palsy study is unique due to the magnitude of patients included, a follow-up period of 1 year, a low percentage of patients lost to follow-up, and use of the sensitive regional and weighted Sunnybrook scale. It was also summarized with a low risk of bias in the Cochrane report by Lockhart et al. All these factors favor the implementation of a clinical predictive study concerning nonrecovery from Bell’s palsy. The Sunnybrook facial grading system has not previously...
been described as a prognostic tool in Bell’s palsy. The aim of the present study was to develop a prognostic model using clinical parameters, including the Sunnybrook scale, to find patients with a poor prognosis as early as possible in the disease.

MATERIALS AND METHODS

Computerized data were collected from the Scandinavian Bell’s palsy trial performed from May 2001 to September 2007. The study included screened patients with an acute unilateral peripheral facial palsy from 16 public otorhinolaryngological centers in Sweden and one in Finland. Patients aged 18 to 75 years with onset of palsy within 72 hours were considered for inclusion. Baseline assessment was performed before the start of treatment. The procedure for randomization and blinding to treatment allocation has previously been described. In total, 829 of 1,953 patients (341 women and 488 men; mean age, 42 years) were included in the modified intention-to-treat analysis. Study design was factorial and thereby included four analysis groups. A total of 416 of the patients were given prednisolone, whereas 413 did not receive it; 413 patients received valacyclovir, whereas 416 were not treated with this drug. Follow-up visits were at days 11 to 17, and at 1, 2, 3, 6, and 12 months. If recovery was complete (Sunnybrook = 100) at 2 or 3 months, the next follow-up was at 12 months.

Outcome Measures

Facial function at 12 months was used as the outcome measure in the present study. The Sunnybrook scale was the main grading system to assess facial function. In addition, the House-Brackmann scoring system was also used. Sunnybrook facial grading is a regionally weighted system that includes evaluation of resting symmetry, degree of voluntary movements, and synkinesis to form a composite score from 0 to 100, where 0 is complete paralysis and 100 normal function. The House-Brackmann system ranges from I (normal) to VI (total paralysis). In the present study, complete recovery was defined as Sunnybrook 100, whereas nonrecovery was defined as a Sunnybrook score <70.

Prediction Model

The data set included 829 patients. They were divided randomly into a training data set including 500 patients (60%) and a validation data set of 329 patients (40%) to first develop and then validate a mathematical model.

The Scandinavian Bell’s palsy trial concluded that prednisolone treatment improves restitution of facial function, whereas valacyclovir treatment had no proven effect. Analyses of valacyclovir treatment performed in the present study showed no significance for nonrecovery. These results are therefore not presented.

Training Data Set

Univariate and multivariate logistic regression analyses, with facial function at 12 months as main outcome, were performed using the training data set (n = 500). Potential prognostic factors of outcome were first analyzed by univariate logistic regression analysis that included patient characteristics at baseline: treatment (prednisolone or no prednisolone), gender, time from onset to inclusion, age, side of palsy, and pain. Statistically significant predictors for nonrecovery were then analyzed for their independent predictive values using multivariate logistic regression.

Data from days 11 to 17 that included Sunnybrook scores, pain, deterioration from baseline to days 11 to 17, and previously mentioned baseline variables were then tested in a multivariate model.

Finally, a third multivariate model was constructed. This included the variables described for the baseline, days 11 to 17 models, and the Sunnybrook scores at 1 month.

For each variable, transformations to satisfy the linearity assumption of the logistic regression models were sought. Missing values were replaced with the last available observation for each patient (i.e., the last-observation-carried-forward [LOCF] method). In addition to LOCF, data obtained by the complete-case analysis method (743 patients) were also analyzed in all models.

Validation Data Set

The logistic regression models obtained with the training data set were also tested on the validation data set (n = 329).

Receiver Operating Characteristics and Area Under the Curve

To evaluate the diagnostic performance of the logistic regression models, receiver operating characteristic (ROC) curves were constructed to discriminate between patients with Sunnybrook scores ≥70 and Sunnybrook scores <70 (nonrecovery) at 12 months. ROC analysis curves show the relationship between sensitivity on the y-axis and specificity on the x-axis for different cutoff levels of test positivity. The area under the ROC curve (AUC), also known as the c-statistic, provides a measure of the overall discriminative power of a model. Values can range from 0.5 (no predictive power, expected by chance alone) to 1 (perfect prediction).

Results are presented with numbers of observations, odds ratios with 95% confidence intervals (CI), AUC (95% CI), sensitivity, and specificity. P values ≤.05 were considered significant. Statistical analyses were carried out using the SAS version 9.2 statistical program (SAS Institute, Cary, NC).

RESULTS

Prediction Model Obtained With the Training Data Set

Table I shows the univariate regression analysis of baseline variables. Among the seven variables, treatment (prednisolone or no prednisolone) (P = .0005), age (P = .04) and Sunnybrook score at baseline (P = .0002) were statistically significant prognostic factors for nonrecovery. Gender, time from onset of palsy to treatment start, side of palsy, and pain at baseline were not significantly correlated with nonrecovery.

The multivariate regression analysis of baseline data is shown in Table II. Treatment, age, and Sunnybrook score at baseline were independent significant predictive factors and were used to create the baseline model.

The multivariate analysis of data obtained from days 11 to 17 is illustrated in Table III. Treatment and age lost their predictive ability when analyzed together with Sunnybrook scores at days 11 to 17. The Sunnybrook score at days 11 to 17 was the only variable that was a significant predictive factor. Multivariate analysis at 1 month also showed the Sunnybrook score as the only significant factor for nonrecovery (Table IV).
Therefore, Sunnybrook scores at these time points during recovery were used to develop the days 11 to 17 and 1-month models for clinical practice.

Figures 1 through 3 show ROC curves for the baseline, days 11 to 17, and 1-month models for the training and validation data sets. The AUC value for nonrecovery in the training data set \( n = 500 \) was 0.74 (sensitivity 0.83, specificity 0.57) for the baseline model and 0.83 (sensitivity 0.81, specificity 0.75) for the days 11 to 17-model. The corresponding value at 1 month was excellent with AUC 0.94 (sensitivity 0.91, specificity 0.85).

Validation of the models in the independent validation data set \( n = 329 \) showed good calibration and discriminative ability for predicting nonrecovery (AUC 0.74 at baseline, 0.87 at days 11 to 17, and 0.94 at 1 month) (Figs. 1–3).

In the present work, the Sunnybrook facial grading system was the main grading system used. This regional and weighted system is more sensitive than the gross House-Brackmann grading scale.\(^{14}\) We performed the same predictive calculations for nonrecovery using the House-Brackmann scale with nonrecovery defined as a House-Brackmann score \( > II \). The results for the two systems were similar (AUC 0.73 at baseline, AUC 0.85 at days 11-17, AUC 0.94 at 1 month for the House-Brackmann calculations, curves not shown).

The predicted probability for a Sunnybrook score \( < 70 \) (nonrecovery) at 12 months was also calculated for the total data set (for patients receiving prednisolone or no prednisolone) using the presented 1-month model. As shown in Figure 4, the risk for nonrecovery predicted at 1 month differs greatly depending on the Sunnybrook score at that time point, both in patients treated and not treated with prednisolone. A Sunnybrook score of 55 at 1 month gives a 5% predicted risk for nonrecovery at 12 months, whereas a Sunnybrook score of 29 at 1 month increases the risk of nonrecovery at 12 months to 30% in patients treated with prednisolone. The corresponding predicted probabilities for nonrecovery using the House-Brackmann scale, where nonrecovery at 12 months was defined as House-Brackmann score \( > II \), are shown in Figure 5.

The predicted results achieved using the LOCF method were compared with the results using the complete-case analysis method \( n = 743 \). Results with the two different methods were similar (results not shown).

**DISCUSSION**

The presented prediction model identifies Bell’s palsy patients at 1 month with a high risk for nonrecovery (Sunnybrook score \( < 70 \)) at 12 months. The accuracy of the ROC model was excellent at 1 month (AUC = 0.94)

### TABLE I.
Univariate Regression Analysis of Baseline Variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>No.</th>
<th>Odds Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>500</td>
<td></td>
<td>.0005</td>
</tr>
<tr>
<td>Prednisolone</td>
<td>250</td>
<td>1 (ref)</td>
<td></td>
</tr>
<tr>
<td>No prednisolone</td>
<td>250</td>
<td>2.95 (1.61–5.41)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>500</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>288</td>
<td>1 (ref)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>212</td>
<td>0.95 (0.55–1.66)</td>
<td></td>
</tr>
<tr>
<td>Time from onset to inclusion, hr</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(&lt; 24)</td>
<td>133</td>
<td>1 (ref)</td>
<td>.08</td>
</tr>
<tr>
<td>(\geq 24)</td>
<td>367</td>
<td>0.60 (0.34–1.07)</td>
<td></td>
</tr>
<tr>
<td>Age, yr</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(&lt; 30)</td>
<td>118</td>
<td>1 (ref)</td>
<td>.04</td>
</tr>
<tr>
<td>30–43</td>
<td>170</td>
<td>2.73 (1.14–6.55)</td>
<td></td>
</tr>
<tr>
<td>44–57</td>
<td>116</td>
<td>1.50 (0.55–4.07)</td>
<td></td>
</tr>
<tr>
<td>(&gt; 57)</td>
<td>96</td>
<td>3.17 (1.25–8.07)</td>
<td></td>
</tr>
<tr>
<td>Side of palsy</td>
<td>500</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>258</td>
<td>1 (ref)</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>242</td>
<td>1.47 (0.85–2.56)</td>
<td></td>
</tr>
<tr>
<td>Pain at inclusion</td>
<td>497</td>
<td></td>
<td>.50</td>
</tr>
<tr>
<td>Yes</td>
<td>245</td>
<td>1 (ref)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>252</td>
<td>0.82 (0.47–1.44)</td>
<td></td>
</tr>
<tr>
<td>Sunnybrook at baseline</td>
<td>499</td>
<td>0.97 (0.95–0.99)</td>
<td>.0002</td>
</tr>
</tbody>
</table>

Main predictors for nonrecovery defined as Sunnybrook \( < 70 \) at 12 months. Data at baseline, training set. CI = confidence interval; ref = reference.

### TABLE II.
Multivariate Regression Analysis of Baseline Data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>No.</th>
<th>Odds Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>499</td>
<td></td>
<td>.0003</td>
</tr>
<tr>
<td>Prednisolone</td>
<td>249</td>
<td>1 (ref)</td>
<td></td>
</tr>
<tr>
<td>No prednisolone</td>
<td>250</td>
<td>3.12 (1.68–5.81)</td>
<td></td>
</tr>
<tr>
<td>Age, yr</td>
<td>499</td>
<td>.058</td>
<td></td>
</tr>
<tr>
<td>(&lt; 30)</td>
<td>117</td>
<td>1 (ref)</td>
<td></td>
</tr>
<tr>
<td>30–43</td>
<td>170</td>
<td>2.82 (1.15–6.93)</td>
<td></td>
</tr>
<tr>
<td>44–57</td>
<td>116</td>
<td>1.37 (0.49–3.82)</td>
<td></td>
</tr>
<tr>
<td>(&gt; 57)</td>
<td>96</td>
<td>2.67 (1.02–6.98)</td>
<td></td>
</tr>
<tr>
<td>Sunnybrook at baseline</td>
<td>499</td>
<td>0.97 (0.95–0.99)</td>
<td>.0002</td>
</tr>
</tbody>
</table>

Main predictors for nonrecovery defined as Sunnybrook \( < 70 \) at 12 months. Data at baseline, training set. CI = confidence interval; ref = reference.

### TABLE III.
Multivariate Analysis of Data Obtained From Days 11 to 17.

<table>
<thead>
<tr>
<th>Variable</th>
<th>No.</th>
<th>Odds Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunnybrook at days 11 to 17</td>
<td>499</td>
<td>0.94 (0.92–0.96)</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Main predictors for nonrecovery defined as Sunnybrook \( < 70 \) at 12 months. Data at days 11 to 17, training set. CI = confidence interval.

### TABLE IV.
Multivariate Analysis of Data Obtained at 1 Month.

<table>
<thead>
<tr>
<th>Variable</th>
<th>No.</th>
<th>Odds Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunnybrook at 1 month</td>
<td>500</td>
<td>0.92 (0.91–0.94)</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Main predictors for nonrecovery defined as Sunnybrook \( < 70 \) at 12 months. Data at 1 month, training set. CI = confidence interval.
and the main predictor for nonrecovery was the Sunnybrook score at 1 month. Our model is validated and simple to use for early identification of patients with a high risk for poor outcome, and it may also be used for the planning and need for follow-up visits.

The predictive value of Sunnybrook grading at baseline was weaker than at days 11 to 17 and at 1 month, which is reflected by the AUC values for the prediction model (AUC = 0.74 at baseline, 0.83 at days 11 to 17, and 0.94 at 1 month). The improvement in predictive capability over time may be explained by the palsy being unstable in its early stage and then stabilizing.

This is the first clinical study in which facial function at 1 month, graded with Sunnybrook, has been evaluated and shown to be a reliable prognostic factor. Our results show that 1 month is an optimal time point...
for clinical prediction of nonrecovery in Bell’s palsy. This is in accordance with the reported findings of Ikeda et al., who found that the House-Brackmann grading at 1 month had a statistical significance for the prognosis of facial paralysis. Our findings of a high predictive capacity of Sunnybrook at 1 month are, however, in contrast to those of Takemoto and coworkers. They reported that Yanagihara grading at baseline as well as the worst grading score during recovery are not significant predictable variables. These contradictive results may be due to the difference in grading systems used (Yanagihara is not weighted), fewer patients included in their study, the time point for the predictive evaluation, and that the definition of nonrecovery was milder in their study.

The minimal response level on electroneurography has traditionally been used to predict recovery in Bell’s palsy patients. Based on the results of multivariate analysis and ROC curves in their study of 142 patients with Bell’s palsy and 26 with Ramsay Hunt syndrome, Takemoto et al. stated that electroneurography was the most effective factor for predicting prognosis in peripheral facial palsy. In general, a response of l<10% indicates a poor prognosis. Electroneurography, however, is not always reliable in predicting outcome in Bell’s palsy and access to this method is limited. This implies a need for clinical factors that can accurately predict prognosis.

The influence of age on outcome has been reported to be highly significant. In the work of Peitersen, only 36% of patients above the age of 60 years experienced the return of normal function. Adour and Wingerd described a less-favorable outcome for patients aged over 60 years. In the work of Ushio et al., age was reported to be a significant factor for the final outcome, but not for time to maximal recovery. In other studies, however, age was not among the most important variables in predicting prognosis of Bell’s palsy. The latter is in agreement with our study, in which age was not a major prognostic factor.

Some of the patients with inadequate improvement of facial function will require plastic reconstructive intervention to achieve better facial symmetry. Today, surgical nerve crossover techniques can be used to improve facial function. This surgery is best carried out before the muscles on the paralyzed side become atrophied. The time interval when surgery needs to be performed is preferably within 9 months after onset of palsy. It is therefore important to predict early those patients who risk suffering sequelae and to identify candidates for surgery. Our prognostic model and the risk curve obtained may, therefore, be an additional tool to identify these patients.

The present study has some limitations. The Scandinavian Bell’s palsy study included 829 patients. Of these, 719 had their facial function assessed at 1 month. In our prediction model, we used imputed data for missing values. Another shortcoming is that the early stage assessments were often performed by less-experienced doctors, which might have influenced the results. A more descriptive and precise version of the Sunnybrook scale has recently been suggested, but this newer scale had not been presented when the Scandinavian Bell’s palsy study was performed.

CONCLUSION

The present study shows that Sunnybrook grading at 1 month accurately predicts nonrecovery at 12 months in Bell’s palsy. A simple-to-use risk curve for nonrecovery at 12 months is presented. Our model may be an additional tool in predicting nonrecovery in the clinical setting and can be used to select patients for further neurophysiological testing.

Acknowledgments

The authors wish to thank Anna Stjernquist-Desatnik, Anne Pitkäranta, Mervi Kanerva, Sara Axelsson, and Per Hanner in the Scandinavian Bell’s palsy study group. We also thank Matti Annikko for encouraging support, Lars Berglund (Biostatistician, Uppsala Clinical Research Centre) for developing the database.

BIBLIOGRAPHY


