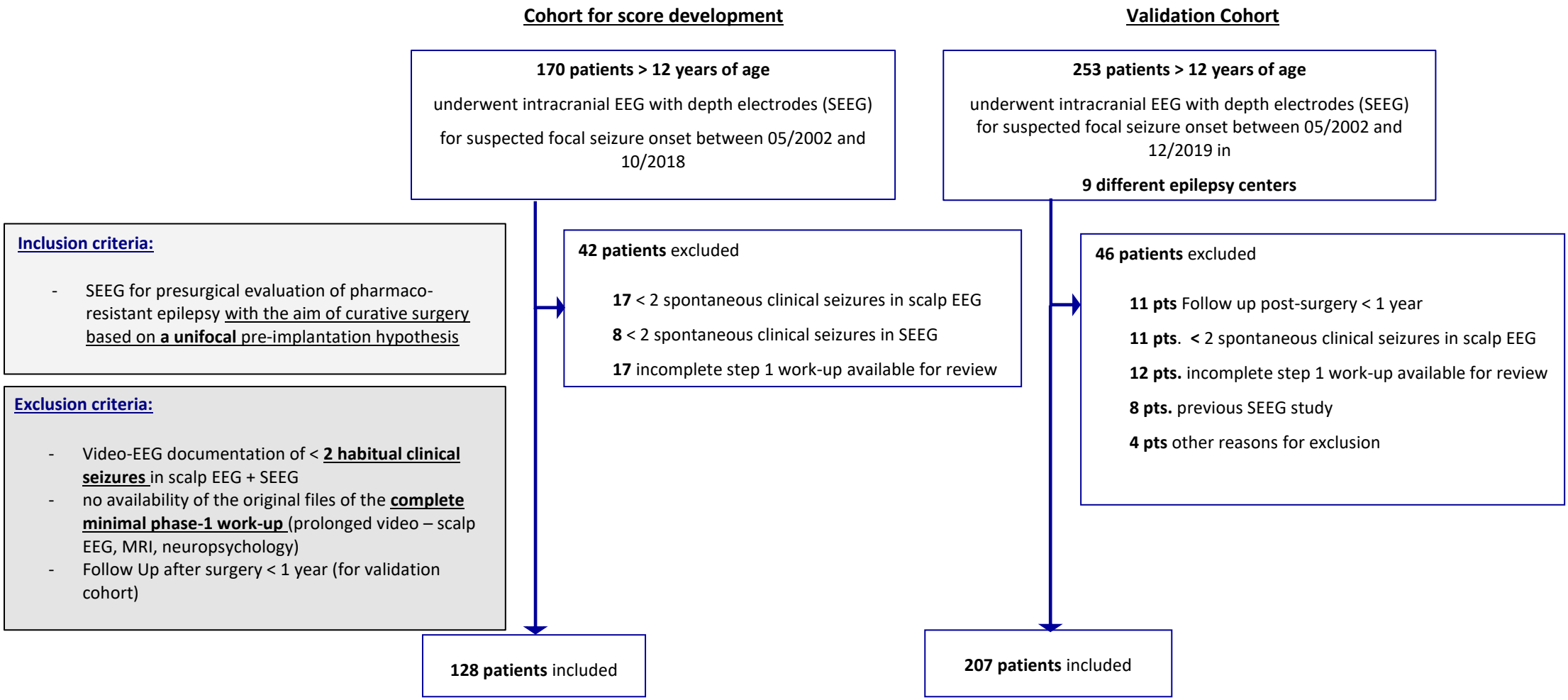
Supplemental Online Content

Astner-Rohracher A, Zimmermann G, Avigdor T, et al. Development and validation of the 5-SENSE score to predict focality of the seizure-onset zone as assessed by stereoelectroencephalography. *JAMA Neurol*. Published online December 6, 2021. doi:10.1001/jamaneurol.2021.4405

This supplemental material has been provided by the authors to give readers additional information about their work.

eFigure. Flow Chart of Patient Selection of cohort for score development and validation cohort



eAppendix. Statistical Approaches

Logistic regression:

The variable selection using logistic regression models was conducted as follows: We considered all subsets containing 5 predictors from the initial list of 14 variables and chose the Akaike information criterion (AIC)[1] as the objective function, which had to be minimized. The AIC is a measure routinely used in model selection to quantify the tradeoff between model fit and complexity. In order to improve the reliability of our findings, and to account for potential uncertainties in the model selection process, we also considered the "top 5" models with respect to AIC, and added the corresponding variables to the list of selected predictors. As a further sensitivity analysis, we also examined the "top 5" models with 4 predictors. Doing so, we finally obtained a list of potential predictors, which were subsequently merged with the selected variables from the machine learning approach, as described in the main body of the manuscript.

Machine Learning:

Data was divided into a 2/3 training set and a 1/3 testing set, while missing values in continuous features were categorized. All categorical variables were transformed into binary dummy variables. Four different models were evaluated as possible prediction models: Random Forest, Naïve Bayes, Support vector machine (SVM) and a basic Neural network (NN). All models were computed using the Scikit-Learn (https://scikit-learn.org/stable/index.html) tool box with python 3.6 (https://www.python.org/downloads/release/python-360). Finally, models were tested for robustness using fifty iterations over different random state initializations when training. This was done to avoid using an outlier state which would be not reproducible in the later stages. In general, based on clinical reasoning, we preferred higher specificity over sensitivity, hence influencing the choice of the final model for each algorithm. The best model was chosen for each type of algorithm configuration as the one that maximizes the following metric of a weighted average of 0.4*sensitivity and 0.6 specificity i.e the model that was chosen was the one with highest weighted average score.

For Random Forest, different class weights (0.1-0.9) and numbers of trees (0-100) were assessed, resulting in a final model with 10 trees and class weights of 0.5 for non-focal and focal, with a specificity of 0.80, and a sensitivity of 0.712. For Naïve Bayes, a specificity of 0.92, and a sensitivity of 0.36 was reached with the selected final model, with class weights for non-focal and focal of 0.2 and 0.8 respectively. Note that there is another model that does not maximize the average, but which is more suitable with a class weight for non-focal vs focal (0.8, 0.2 respectively) with a result of

specificity = 0.52 and sensitivity = 0.78. For the final model of the Support Vector Machine approach, a regularization parameter of 1 and a class weight for non-focal vs focal of 0.4 vs. 0.6 was set respectively, resulting in a specificity of 0.88 and sensitivity of 0.42. Note that there is another model that does not maximize the average but is more suitable with a class weight for non-focal vs focal (0.3, 0.7 respectively) with specificity of 0.64 and sensitivity of 0.64. The final model of the Neural Network (NN) algorithm comprised of 8 hidden layers with 16 neurons after assessment by a backpropagation algorithm with a 'relu' activation, resulting in a specificity of 0.84 and a sensitivity of 0.71.

We ranked the results for each model by importance of the parameters and selected the two, in case they were almost equally important three, most important parameters. To estimate the importance of each parameter, independently from other variables, a permutation importance approach [2]was applied for each model. This approach determines the importance of each parameter by assessing the decline of the predictive performance compared to baseline performance by randomly shuffling this parameter.

Parameters identified by the four different machine learning algorithms:

Random Forest: EEG_clinic, iEEGonset_ext_3, lesion_ext_3
Naïve Bayes: N_szTypes, iEEGonset_ext_3

For the second model: EEG_clinic, iEEGonset_ext_3

Support Vector Machine (SVM): EEG_clinic, lesion_ext_3

For the second model: lesion_ext_3, NPSY_loc (a lot less)

Neural Network (NN): EEG_clinic, lesion_ext_3

Score Development

Figure 1 describes the stepwise approach used to develop the score. In step 3, selected predictors were merged into one single list (Figure 1) and the logistic regression model with five variables from that list, which minimized the Akaike information criterion (AIC), was considered as the final model. To adequately reflect the fact that several models had similar AIC values, models in which one of the variables was either dropped or replaced by another variable were also considered. The decision regarding the "final" models was based on clinical and statistical considerations.

Mathematically, the “5-SENSE-Score” is calculated as a predicted probability following standard logistic regression methodology [3]: At first, the model coefficients of each variable, corresponding to the characteristics of an individual patient, are summed up, i.e. if the patient has no lesion on MRI, the value “-2.2636” is put in the first line, if the ictal EEG onset is focal the value “0.8442” is added in the second line, etc.. Moreover, irrespective of the patient, the model-specific constant “-0.3135” is added as a statistical intercept of the model, yielding the value denoted as *sum* in the sequel. As a second step an exponential function is used to calculate the probability of focality: *Probability (focal) = exp(sum) / (exp(sum) + 1).* Finally, to obtain a score value between 0 and 100, the calculated probability is multiplied with 100. If the resulting value exceeds 37.6, the score is pointing to a focal SOZ in SEEG in this patient.

eTable. Demographic data of the patient cohorts used for score development and validation

	Cohort for score development				Validation cohort			
	Overall (n=128)	**Focal (n=48)**	**Non-focal (n= 80)**	**p- value***	**Overall (n=207)**	**Focal (n=111)**	**Non-focal (n= 96)**	**p- value***
Age (years), median [range]	32 [13;58]	31 [13; 57]	32 [13; 59]	p= 0.77	32 [16;70]	33.5 [16; 66]	30 [16; 70]	p=0.13
Sex, f, n (%)	57 (45)	19 (40)	38 (48)	p= 0.46	96 (47)	50 (45)	47 (49)	p= 0.58
Age at onset (years) median [range]	13 [0.3; 59]	14 [2;38]	12.5 [0.3; 59]	p= 0.96	12 [0; 64]	13 [0;64]	10 [0; 55]	p=0.20
Duration of epilepsy (years) median [range]	17 [2;53]	20 [2;53]	16 [2; 53]	p= 0.64	19 [1; 64]	19 [1; 64]	18.5 [2.5; 51]	p= 0.58
Previous epilepsy surgery	19 (15)	8 (17)	11 (14)	p =0.82	49 (24)	16 (14)	33 (34)	p = 0.001
Neuropsychological examination				P=0.16				P=0.13
- Normal	10 (8)	5 (10)	5 (6)		8 (4)	7 (6)	1(1)	
- Localizing deficits	45 (35)	21 (44)	24 (30)		90 (43)	45 (41)	45 (47)	
- Global (non-localizing) deficits	71 (56)	21 (44)	50 (63)		109 (53)	59 (53)	50 (52)	
- Not available	2 (2)	1 (2)	1 (1)					
Etiology, n (%)				p=0.19				
- Malformation of cortical development	51 (40)	20 (32)	31 (39)		48 (23)	23 (21)	25 (26)	P=0.34
- Hippocampal sclerosis	8 (6)	5 (10)	3 (4)		15 (7)	9 (8)	6 (6)	
- Genetic	3 (2)	2 (4)	1 (1)		8 (4)	4 (4)	4 (4)	
- Cerebrovascular	5 (4)	3 (6)	2 (3)		5 (2)	1 (1)	4 (4)	
- Posttraumatic	11 (9)	3 (6)	8 (10)		9 (4)	2 (2)	7 (7)	
- Postinfectious	7 (5)	2 (4)	5 (6)		8 (4)	3 (3)	5 (5)	
- Perinatal	5 (4)	2 (4)	3 (4)		7 (3)	2 (2)	5 (5)	
- Tumor	2 (2)	2 (4)	0 (0)		7 (3)	5 (5)	2 (2)	
- Autoimmune					2 (1)	0 (0)	2 (2)	
- Unknown	36 (28)	9 (19)	27 (33.7)		92 (44)	56 (50)	36 (38)	
- Other					6 (3)	6 (5)	0 (0)	

*For metric/ordinal variables, the t-test for unequal variances (i.e., the Satterthwaite-Smith-Welch approximation) was used. For nominal variables, Fisher's Exact test was applied.

eReferences

1. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974;19:716-723.
2. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32. doi:10.1023/A:1010933404324.
3. Agresti A. *Categorical data analysis.* 3rd ed. New Jersey: John Wiley & Sons, Ltd; 2013.