

RESEARCH ARTICLE

Open Access



Validation of a prediction model for long-term outcome of aphasia after stroke

Femke Nouwens^{1*}, Evy G. Visch-Brink¹, Hanane El Hachoui², Hester F. Lingsma³, Mieke W. M. E. van de Sandt-Koenderman^{2,4}, Diederik W. J. Dippel¹, Peter J. Koudstaal¹ and Lonneke M. L. de Lau^{1,5}

Abstract

Background: About 30% of stroke patients suffer from aphasia. As aphasia strongly affects daily life, most patients request a prediction of outcome of their language function. Prognostic models provide predictions of outcome, but external validation is essential before models can be used in clinical practice. We aim to externally validate the prognostic model from the Sequential Prognostic Evaluation of Aphasia after stroke (SPEAK-model) for predicting the long-term outcome of aphasia caused by stroke.

Methods: We used data from the Rotterdam Aphasia Therapy Study – 3 (RATS-3), a multicenter RCT with inclusion criteria similar to SPEAK, an observational prospective study. Baseline assessment in SPEAK was four days after stroke and in RATS-3 eight days. Outcome of the SPEAK-model was the Aphasia Severity Rating Scale (ASRS) at 1 year, dichotomized into good (ASRS-score of 4 or 5) and poor outcome (ASRS-score < 4). In RATS-3, ASRS-scores at one year were not available, but we could use six month ASRS-scores as outcome. Model performance was assessed with calibration and discrimination.

Results: We included 131 stroke patients with first-ever aphasia. At six months, 86 of 124 (68%) had a good outcome, whereas the model predicted 88%. Discrimination of the model was good with an area under the receiver operation characteristic curve of 0.87 (95%CI: 0.81–0.94), but calibration was unsatisfactory. The model overestimated the probability of good outcome (calibration-in-the-large $a = -1.98$) and the effect of the predictors was weaker in the validation data than in the derivation data (calibration slope $\beta = 0.88$). We therefore recalibrated the model to predict good outcome at six months.

Conclusion: The original model, renamed SPEAK-12, has good discriminative properties, but needs further external validation. After additional external validation, the updated SPEAK-model, SPEAK-6, may be used in daily practice to discriminate between patients with good and patients with poor outcome of aphasia at six months after stroke.

Trial registration: RATS-3 was registered on January 13th 2012 in the Netherlands Trial Register: [NTR3271](https://www.trialregister.nl/trial/1271). SPEAK was not listed in a trial registry.

Keywords: Aphasia, Stroke, Prognostic model, Validation, Outcome

* Correspondence: f.nouwens@erasmusmc.nl

¹Department of Neurology, Erasmus MC University Medical Center, room Nb-324, PO Box 2040, 3000, CA, Rotterdam, the Netherlands
Full list of author information is available at the end of the article



Background

Aphasia occurs in approximately 30% of stroke patients and has a strong impact on everyday communication and daily functioning [1, 2]. Shortly after stroke, patients and their family are faced with major uncertainties regarding recovery of communication. Consequently, there is a need for individual estimation of the expected outcome. Adequate personal prognosis may also contribute to optimizing individual care, which is important as medical and paramedical care becomes increasingly personalized [3]. Furthermore, predictions of outcome may corroborate rationing of care, in order to better distribute limited resources. Prediction of post-stroke aphasia outcome is often based on models that consist of determinants identified in a single dataset, e.g. age, sex, aphasia severity and subtype; site, size and type of the lesion; vascular risk factors and stroke severity [4–11]. Before a model can be used in daily practice, it should be externally validated [3, 12]. This means that the generalizability of a model is assessed in different cohorts with more recent recruitment (temporal validation), from other institutions (geographical validation), and by different researchers [3]. To our knowledge, none of the few available prognostic models predicting language outcome has been externally validated [13–16].

Previously, our group has constructed a prognostic model for the outcome of aphasia due to stroke. The model was derived from the dataset of the Sequential Prognostic Evaluation of Aphasia after stroke (SPEAK) study, and performed well [13]. The aim of the current study was to externally validate the SPEAK-model in an independent, yet comparable cohort of stroke patients with aphasia.

Methods

The SPEAK-model

SPEAK was an observational prospective study in 147 patients with aphasia due to stroke conducted between 2007 and 2009 in the Netherlands [13]. Demographic, stroke-related and linguistic characteristics of 130 participants, collected within six days of stroke, were used to construct a model predicting good aphasia outcome one year after stroke, defined by a score of 4 or 5 on the Aphasia Severity Rating Scale (ASRS) from the Boston Diagnostic Aphasia Examination [17]. This scale is used for rating communicative ability in spontaneous speech. The ScreeLing, an aphasia screening test designed to assess the core linguistic components semantics, phonology and syntax in the acute phase after onset, was also included in the model [18–20]. For detailed methods, results and discussion we refer to the original paper [13]. The final SPEAK-model contained six baseline variables: ScreeLing Phonology score, Barthel Index score, age, level of education (high/low), infarction with a cardio-embolic source (yes/no) and intracerebral hemorrhage (yes/no) (Online

Additional file 1, Box A1). This model explained 55.7% of the variance in the dataset. Internal validity of the model was good, with an AUC (Area Under the Curve, where the curve is the Receiver Operation Characteristic (ROC) curve) of 0.89 [13].

Validation

For external validation of the SPEAK-model we used data from the Rotterdam Aphasia Therapy Study (RATS) – 3, a randomized controlled trial (RCT) studying the efficacy of early initiated intensive cognitive-linguistic treatment for aphasia due to stroke, conducted between 2012 and 2014 [21, 22]. RATS-3 was approved by an independent medical ethical review board. Details about the study design, methods and results have been reported elsewhere and a summary will be provided below [21, 22].

Participants and recruitment

A total of 23 hospitals and 66 neurorehabilitation institutions across the Netherlands participated in RATS-3. The majority of participating institutions and local investigators (90%) differed from those involved in SPEAK. In- and exclusion criteria for both studies are presented in Table 1.

Prognostic variables

Patients with aphasia due to stroke were included in RATS-3 within 2 weeks of stroke. At inclusion, the following baseline variables were recorded: age, sex, education level, stroke type (cerebral infarction or intracerebral hemorrhage), ischemic stroke subtype (with or without a cardio-embolic source). Level of independence was estimated with the Barthel Index, a questionnaire containing ten items about activities of daily life [23]. All participants were tested with the ScreeLing to detect potential deficits in the basic linguistic components [19, 24]. Spontaneous speech samples were collected with semi-standardized interviews according to the Aachen Aphasia Test-manual [25]. Aphasia severity was assessed by scoring the spontaneous speech samples with the ASRS.

Outcome

In SPEAK, ASRS-scores were used to assess aphasia outcome [17]. This six point scale is used to rate spontaneous speech and ranges from 0: “No usable speech or auditory comprehension” to 5: “Minimal discernible speech handicaps; the patient may have subjective difficulties which are not apparent to the listener”. The SPEAK-model predicts the occurrence of ‘good outcome’, i.e. an ASRS-score of 4 or 5 after 1 year. In RATS-3 follow-up was at 4 weeks, 3 and 6 months after randomization. ASRS-scores from the RATS-3 cohort at 6 months after randomization were used as outcome in the analysis, as this was closest in time to the original model.

Table 1 In- and exclusion criteria for participants in RATS-3 and in the SPEAK cohort

	RATS-3	SPEAK
Inclusion:	First-ever aphasia due to stroke Aphasia ascertained by a speech and language therapist using the 36-item Token Test ^a and/or a score < 5 on the ASRS Testable with the ScreeLing Within two weeks of stroke onset Age between 18 and 85 Language near-native Dutch A life expectancy of >six months Able to tolerate intensive treatment	First-ever aphasia due to stroke Aphasia ascertained by a neurologist and a speech and language therapist A score below the cut-off point of the Token Test and/or the ScreeLing Within two to six days of stroke onset Adult Language near-native Dutch
Exclusion:	A subarachnoid or subdural hemorrhage Success or feasibility of intensive language treatment was severely threatened by: <ul style="list-style-type: none"> - severe dysarthria - premorbid dementia - illiteracy - severe developmental dyslexia - severe visual perceptual disorders - recent psychiatric history 	Presence of one of the following criteria: <ul style="list-style-type: none"> - severe dysarthria - pre-stroke dementia (suspected or confirmed) - illiteracy - developmental dyslexia - severe perceptual disorders of vision or hearing - psychiatric history

^aDe Renzi, E, Faglioni, P. Normative data and screening power of a shortened version of the Token Test. Cortex 1978;14:41–49

Statistical analyses

Outcome in the RATS-3 cohort was divided in good (ASRS-score of 4 or 5) or poor (ASRS-score < 4). To validate the SPEAK-model we assessed discrimination and calibration [3, 12, 26–28]. For both analyses predicted probability of a good outcome was calculated using the SPEAK-model (Online Additional file 1, Box A1).

Discriminative properties of the model were summarized with the *c* index, similar to the AUC. Good discrimination means that the model is able to reliably distinguish patients with good aphasia outcome from those with poor outcome.

We assessed the calibration properties of the model by studying to what extent the predicted probability of aphasia outcome corresponded with the observed outcome. A calibration plot was constructed by ordering the predicted probabilities of good aphasia outcome ascendingly and forming five equally large groups. Per group, the mean probability of a good outcome at 6 months was calculated, resulting in five predicted risk-groups. Subsequently, in each risk-group, proportions were calculated of participants with an observed good outcome. These proportions were plotted against the mean probability of a good outcome predicted by the SPEAK-model. Outcomes of the linear predictor *y*, calculated with the SPEAK-model, were used to fit a logistic regression model predicting the dichotomous outcome of

good versus poor outcome to assess calibration-in-the-large and the calibration slope. If calibration of a model is optimal, the calibration-in-the-large α equals 0 and the calibration slope β equals 1. In case of insufficient calibration we will recalculate the prognostic model by adjusting the intercept.

Handling of missing data

For participants with missing outcome scores at 6 months, scores at 3 months after randomization were used. If no scores were available at 3 months, patients were excluded. Missing data for the other variables were imputed using simple imputation: for binary and categorical variables the mode was imputed and means were used for continuous variables.

Results

No outcome data at 6 months were available in 28 of 153 participants, and one participant was excluded because aphasia was later found to be caused by a brain tumor. Reasons for missing outcome data were death (*n* = 7), serious illness (*n* = 4), refusal (*n* = 16) and emigration abroad (*n* = 1). Of these 28 patients, 21 participants were excluded because outcome at 3 months was also not available. For 7 participants we used ASRS-scores at 3 months to impute missing values at 6 months. Baseline data of patients in the validation sample (*n* = 131), as

well as those from the SPEAK cohort ($n = 147$) are provided in Table 2. Groups differed slightly with respect to the baseline variables sex, level of education, type of stroke and aphasia severity.

In the derivation SPEAK cohort ($n = 130$), 11% of the patients had an ASRS-score of 4 or 5 at baseline (4 days after stroke) and 78% had a good outcome after 1 year. In the RATS-3 cohort we found a proportion of 21% with a score of 4 or 5 at baseline (8 days after inclusion) and 68% at 6 months. This is comparable to the 74% in SPEAK at six months. The course of ASRS-scores in the RATS-3 and SPEAK cohort over time is presented in Fig. 1.

Discrimination of the SPEAK-model was good, with an AUC of 0.87 (95% confidence interval: 0.81 to 0.94). In Fig. 2a, the grey line depicts calibration of a hypothetically perfect model and the 5 dots represent calibration values in the five subgroups of patients, ordered by increasing predicted probabilities and plotted against the actual proportions of good outcome. The mean predicted probability of good aphasia outcome at 1 year was 88%, while the observed percentage was 68%, but this was measured at 6

months. The SPEAK-model was too optimistic in predicting good aphasia outcome, with calibration-in-the-large of $\alpha = -1.98$. The calibration slope of $\beta = 0.88$ indicated that the predictor effects were slightly weaker in the validation data than in the derivation data.

As Fig. 1 shows that there is still improvement after 6 months, we assume that the poor calibration-in-the-large is at least partly due to the different timing of the outcome measurement; 6 months versus 1 year. Thus, we updated the SPEAK-model to predict outcome at 6 months instead of 1 year by adapting the intercept (Online Additional file 1, Box A2). After revising the SPEAK-model, the calibration slope remained $\beta = 0.88$, but calibration-in-the-large improved considerably: $\alpha = -0.24$ (Fig. 2b). We suggest renaming the original SPEAK-model predicting outcome at 1 year after stroke into SPEAK-12 and naming the updated model SPEAK-6.

Discussion

We aimed to externally validate the published SPEAK-model for the long-term prognosis of aphasia due

Table 2 Baseline model parameters of participants in the original SPEAK cohort and in RATS-3

	SPEAK cohort ($n = 147$) Derivation cohort	RATS-3 cohort ($n = 131$) Validation cohort
Age, mean (SD), in years	67 (15)	65 (12)
Sex, n (%female)	78 (53%)	56 (43%)
Level of education, n (%)		
High [■]	55 (42%)	60 (46%)
Low [▲]	74 (57%)	71 (54%)
Unknown [◆]	2 (2%)	0
Type of stroke, n (%)		
Non-cardio-embolic infarction	84 (57%)	81 (62%)
Cardio-embolic infarction	42 (29%)	23 (18%)
Intracerebral hemorrhage	21 (14%)	24 (18%)
Unknown [◆]	0	3 (2%)
Time since onset to inclusion, mean (range), in days	4 (2–6)	8 (1–18)
Barthel Index, median (IQR) [◆]	15 (7.75–20)	16 (6–20)
ScreeLing Phonology score, mean (SD) [○]	14 (6)	15 (6.5)
ASRS-scores at baseline, n (%)		
Score 0	18 (12%)	17 (13%)
Score 1	28 (19%)	21 (16%)
Score 2	33 (22%)	28 (21%)
Score 3	26 (18%)	38 (29%)
Score 4	27 (18%)	27 (21%)
Score 5	3 (2%)	0
Missing	12 (8%)	0

[■] High = senior vocational education, higher education or university

[▲] Low = no/unfinished elementary school, elementary school, unfinished junior secondary vocational education or junior secondary vocational education

[◆] Imputed scores used for analysis: level of education = low; type of stroke = non-cardio-embolic infarction; Barthel Index score = 13 ($n = 14$)

[○] ScreeLing Phonology scores range from 0 to 24

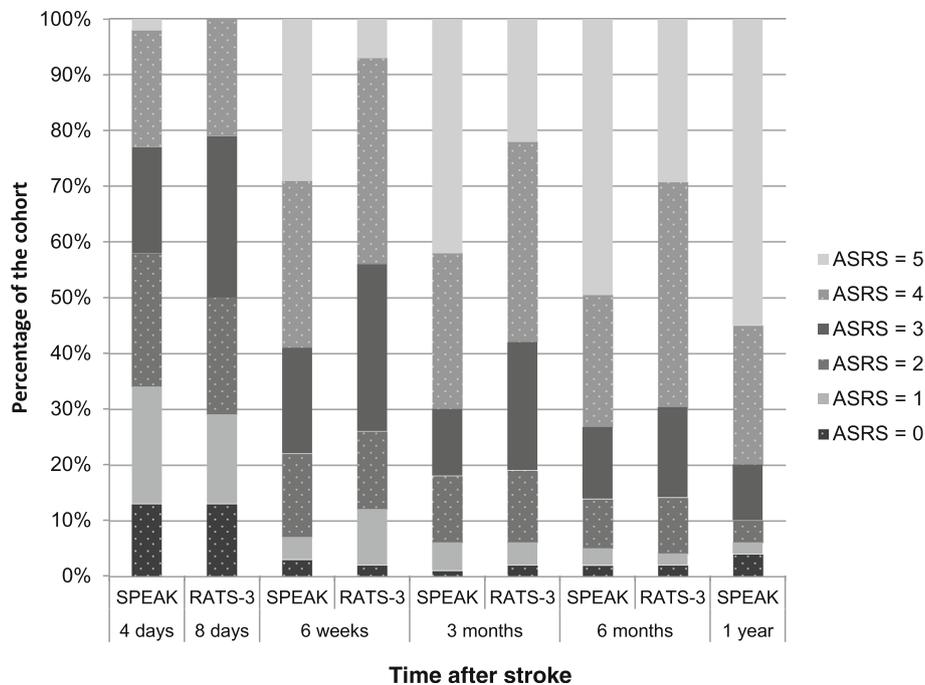


Fig. 1 ASRS-scores over time in SPEAK and RATS-3. ASRS-scores: 5 = minimal discernible speech handicap, some subjective difficulties that are not obvious to the listener; 4 = some obvious loss of fluency in speech or facility of comprehension, without significant limitation in ideas expressed or form of expression; 3 = able to discuss almost all everyday problems with little or no assistance, reduction of speech and/or comprehension; 2 = conversation about familiar topics is possible with help from the listener, there are frequent failures to convey an idea; 1 = all communication is through fragmentary expression, great need for inference, questioning and guessing by listener, limited information may be conveyed; 0 = no usable speech or auditory comprehension

to stroke using data from an independent cohort of stroke patients with aphasia, RATS-3. The SPEAK-model performed very well in terms of discriminating between good (ASRS 4 or 5) and poor (ASRS < 4) outcome. However, calibration was suboptimal, as it was overoptimistic in predicting good aphasia outcome, partly due to the difference in timing of the outcome which was 1 year in SPEAK and 6 months in RATS-3. Therefore, we proposed an updated version of the SPEAK-model for the prediction of outcome at 6 months.

Prognostic models are used in clinical practice to predict possible outcomes or risks of acquiring certain diseases. To our knowledge, apart from the SPEAK-model, only three other models to predict outcome of aphasia after stroke have been published [14–16]. One logistic regression model predicting early clinical improvement in stroke patients with aphasia was constructed based on findings from CT-angiography and CT-perfusion [15]. Clinical applicability of this model is limited, as these detailed CT-data are rarely available in daily practice. Another logistic regression model addressed the effect of speech and language treatment (SLT) on communication outcomes [14]. The authors found that the amount of SLT, added to baseline aphasia severity and baseline stroke disability significantly affected communication 4 to 5 weeks after stroke. Baseline variables

were recorded within 2 weeks of stroke. Recently, a model was published predicting everyday communication ability (Amsterdam-Nijmegen Everyday Language Test; ANELT) at discharge from inpatient rehabilitation based on ScreeLing Phonology and ANELT-scores at rehabilitation admission [16]. These models predict outcome of aphasia recovery only in patients treated with SLT, but do not predict outcome before treatment is initiated. Furthermore, in both studies the cohort included only patients eligible for intensive treatment.

For a prognostic model to be valid and reliable, it is important to evaluate the clinical applicability and generalizability of the model [29]. Inclusion criteria in SPEAK and RATS-3 were not strict, so that both cohorts can be considered representative of acute stroke patients with aphasia in general. The SPEAK-model is valuable for predicting aphasia outcome early after stroke in clinical practice as it includes easily available baseline variables [13]. It requires only the Barthel Index score and the ScreeLing Phonology score to be collected outside clinical routine. The Barthel Index is commonly assessed in the acute phase, allowing for application of this model without much effort [30].

Our study is the first to validate a model for the prognosis of aphasia outcome in an independent cohort. Determining

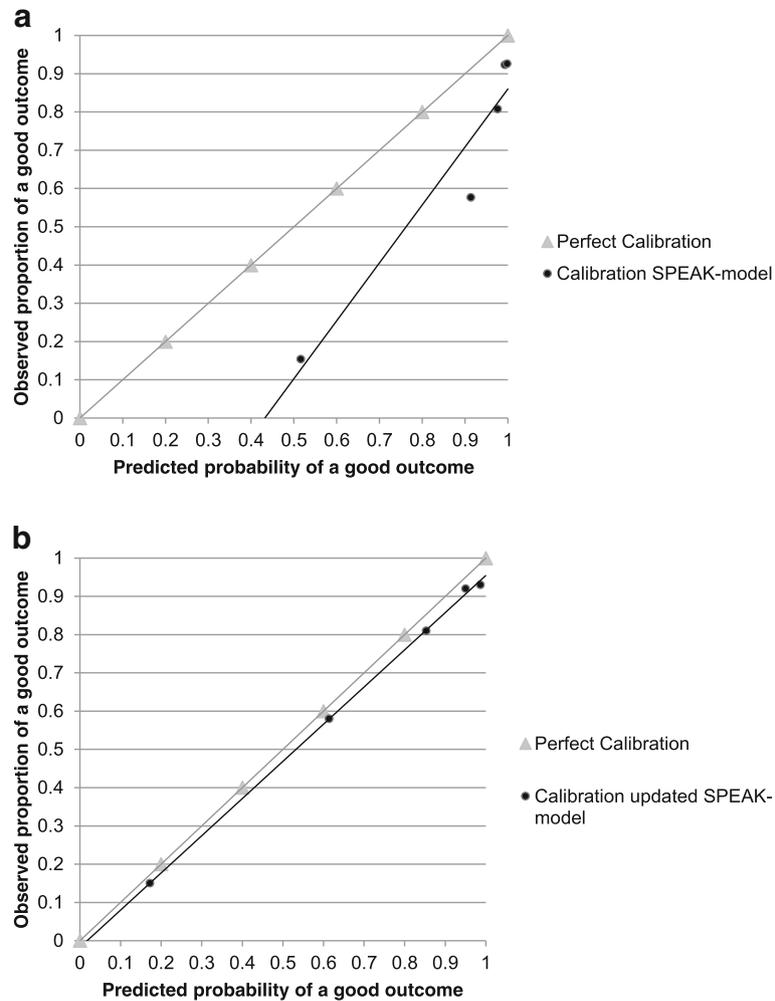


Fig. 2 Calibration plots of the SPEAK-model and updated SPEAK-model. **a** Calibration plot of the original SPEAK-model, SPEAK-12. **b** Calibration plot of the updated SPEAK-model, SPEAK-6

whether a model generalizes well to patients other than those in the derivation cohort, is crucial for the application of that model in daily practice [12, 26, 27, 29]. We found that the SPEAK-model is able to adequately distinguish stroke patients with aphasia who will recover well with respect to functional verbal communication from patients who will not. The model appears less accurate when it comes to the comparison of predicted and actual good outcome.

A first possible explanation may be the different intervention in the two studies. In SPEAK, patients received usual care and researchers did not interfere with the treatment provided. In RATS-3, treatment was strictly regulated, as in this RCT patients were randomly allocated to 4 weeks of either intensive cognitive-linguistic treatment or no treatment, starting within 2 weeks after stroke. After this period both groups received usual care, as in SPEAK. In RATS-3 we found no effect of this early intervention and both intervention groups scored equally on all

outcomes. Thus, we believe treatment does not explain the poor calibration.

Second, there was a difference between SPEAK and RATS-3 with respect to the interval between stroke onset and inclusion of patients. In SPEAK, patients were included on average 4 days after onset and in RATS-3 after 8 days. This seemingly small difference might in fact have caused substantial differences in the prognostic effect of the baseline ScreeLing and Barthel Index scores. Recovery can occur rapidly early after stroke, as was shown in the SPEAK cohort, with a statistically significant improvement on the ScreeLing Phonology score between the first and second week after stroke [31]. Hence, these predictors might have different effects in the RATS-3 cohort, as represented in the suboptimal calibration slope.

Third and most importantly, calibration is likely to have been influenced by a different follow up duration, which was 6 months in RATS-3 versus 1 year

in SPEAK. In SPEAK, ASRS-scores improved significantly up to 6 months after aphasia onset, but no significant improvement was found between 6 and 12 months [31]. We used this finding for the design of the present study to justify the earlier time-point for the outcome in RATS-3. Although in SPEAK no statistically significant improvement in ASRS-scores was found between 6 and 12 months after stroke, some improvement still occurred [31]. Of the participants from SPEAK 74% had an ASRS-score of 4 or 5 at 6 months after stroke, which is fairly similar in RATS-3 at that time-point (68%). It is likely that calibration would have been better if the outcome was determined at 12 months in the RATS-3 cohort, because of the small, but apparent recovery between 6 and 12 months after stroke.

We therefore suggest an updated version, SPEAK-6, to predict outcome at 6 months. More extensive updating could imply refitting the models to the new dataset, to obtain new model coefficients [32–34]. However, as the model discrimination was good, we updated only the intercept to make the model applicable to predict outcome at 6 months, when the average probability of a good outcome is lower than at 1 year. We recommend that the updated SPEAK-6 is validated in the future in new independent datasets.

This study shows again that the external validity of prognostic models in new settings should always be carefully assessed. However, it should also be noticed that perfect calibration might in fact be impossible, as it implies that a model perfectly predicts outcome for all patients [35].

Strengths and limitations

The major limitation of this validation study is the difference in time post onset at which predictor and outcome data were collected. Strength is that the RATS-3 and SPEAK cohorts are comparable, due to similar inclusion criteria. However, whereas participation in SPEAK merely involved periodic language evaluations, RATS-3 was an intervention trial, with either early intensive treatment or no early treatment. Due to these experimental interventions many patients refused participation. Also, selection criteria for RATS-3 were slightly stricter than in SPEAK regarding the potential to receive early intensive treatment. Consequently, the SPEAK and RATS-3 cohorts might represent slightly different populations of stroke patients with aphasia, albeit closely related [26]. Therefore, as in all clinical trials, one must be careful in generalizing the results to all stroke patients with aphasia [36].

Although both the derivation cohort and the validation cohort consist of well over a hundred participants, sample sizes may be considered rather small for adequate modelling [28, 36]. This is reflected in the slight

imbalance of baseline characteristics between both study cohorts. This imbalance may underpin the necessity of larger sample sizes to better reflect the population of stroke patients with aphasia. Furthermore, in both cohorts a fairly large proportion of patients with mild aphasia (baseline ASRS >3) was included. Although this was a reflection of the population, it may have influenced the model, as patients with mild aphasia are known to often fully recover [6, 37].

A much debated issue is the potential lack of sensitivity of rating scales for analyses of spontaneous speech in aphasia [38]. In the current study, we dichotomized outcome, further reducing sensitivity. It can be argued that the definition of “good outcome” with an ASRS of 4 or 5 is somewhat optimistic. A score of 4, or sometimes even 5, does not imply full recovery. Patients with a score of 4 still experience difficulties with word finding or formulating thoughts into language.

The ScreeLing is currently only available in Dutch, which severely limits the applicability of the prediction model. However, adaptation to other languages should not be very complicated as the ScreeLing Phonology subscale contains well-known tasks to measure phonological processing, e.g. repetition, discrimination of minimal pairs, and phoneme/grapheme conversion [20]. Including linguistic functioning, as a possible predictor in prognostic models seems essential, as linguistic functioning, in particular phonology, appears to be a better predictor than overall aphasia severity [13, 16, 39]. At the moment the ScreeLing is being translated into English, German and Spanish, and more languages are to come.

The Barthel Index was included in the original model as a measure of overall stroke severity. Although its reliability is good, it may be debated whether the Barthel Index score accurately reflects stroke severity, as the score is influenced by factors related to health care management choices, such as receiving a urinary catheter or feeding tube on intensive care units. Despite its shortcomings the Barthel Index was found to be a valid outcome measure for stroke trials [40].

Finally, the RATS-3 database contained several missing values. Of the participants who refused evaluation at 6 months, three had fully recovered, which may have introduced a slight bias. Missing values for other variables in the model mostly resulted from inconsistencies in reporting the scores. We used generally accepted methods for imputation of the data and for most variables few data were missing (<5%) [27]. For the Barthel Index 10% had to be imputed, which is a fairly large proportion. There were no clear reasons for these missing values, other than clinicians sometimes just forgot to fill out the score form, which in our view justifies imputation.

Conclusion

The original SPEAK-model, renamed SPEAK-12, performs well in predicting language outcome after 1 year in patients with aphasia due to stroke. As calibration was initially unsatisfactory, we propose an updated version of SPEAK-12 for the prediction of the probability of good language outcome at 6 months: SPEAK-6. Further external validation of SPEAK-12 and SPEAK-6 is recommended. Special attention should be given to timing, as time after stroke onset at which predictors and outcome data are collected appears crucial for adequate model validation. Our results show that SPEAK-6 may be used in daily practice to discriminate between stroke patients with good and patients with poor language outcome at 6 months after stroke.

Additional file

Additional file 1: Appendix 1. The SPEAK-model. (DOCX 40 kb)

Abbreviations

ASRS: Aphasia Severity Rating Scale; AUC: Area Under the ROC Curve; RATS-3: Rotterdam Aphasia Therapy Study – 3; RCT: Randomized Controlled Trial; ROC: Receiver Operation Characteristic; SPEAK: Sequential Prognostic Evaluation of Aphasia after stroke

Funding

SPEAK was funded by the Netherlands Organization for Scientific Research (NWO; grant number 017.002.083). RATS-3 was funded by a fellowship from the Brain Foundation Netherlands granted to LL (grant number: 2011(1)-20). The funding bodies had no role in the design of the study; collection, analysis, and interpretation of data; and in writing the manuscript.

Availability of data and materials

The anonymized datasets used and analyzed for the current study supporting the conclusions in the current article are available from the corresponding author on reasonable request.

Authors' contributions

FN, LL, EV and MS designed the study. FN collected and prepared data for analyses. HL and FN performed statistical analyses. The first version of the manuscript was drafted by FN, and further versions were edited by LL, EV, MS, HL, HH, DD and PK. All authors have critically revised the manuscript and have read and approved the final manuscript.

Ethics approval and consent to participate

SPEAK (MEC-2006-246) and RATS-3 (MEC-2005-347) were approved by the Medical Ethical Committee of Erasmus MC University Medical Center. According to Dutch regulation monitored by the Central Committee on Research Involving Human Subjects local sites formally approved their participation: Bronovo Hospital The Hague, Canisius-Wilhelmina Hospital Nijmegen, Deventer Hospital, Gelre Hospital Zutphen, Haga Hospital The Hague, Kennemer Gasthuis Haarlem, Maasstad Hospital Rotterdam, MCH The Hague, St. Elisabeth Tilburg, Tergooi Hospitals Laren and Blaricum, TweeSteden Hospital Tilburg, UMC St. Radboud Nijmegen, Bernhoven Hospital Oss, Rijnstate Hospital Zevenaar, Diaconessenhuis Meppel, Haven Hospital Rotterdam, Sint Franciscus Gasthuis Rotterdam, Ikazia Hospital Rotterdam, Vlietland Hospital Schiedam, IJsselland Hospital Capelle aan den IJssel, Reinier de Graaf Gasthuis Delft, VUMC Amsterdam, Beatrix Hospital Gorinchem, Amphia Hospital Breda, Onze Lieve Vrouwe Gasthuis Amsterdam, Sint Lucas Andreas Hospital Amsterdam, Catharina Hospital Eindhoven, Franciscus Hospital Roosendaal, Isala Klinieken Zwolle, Jeroen Bosch Hospital Den Bosch, Erasmus MC Rotterdam, Trappenberg Huizen, Leijpark Rehabilitation Tilburg, Tolbrug Rehabilitation Den Bosch, RMC Groot Klimmendaal Ede, Sint Maartenskliniek Nijmegen, Sophia The Hague, Sutfene Warnsveld, Amaris Gooizicht Hilversum, Regina Pacis Arnhem, Gelders Hof Arnhem, Watersteeg Veghel, Den Ooiman Doetinchem, Hazelaar Tilburg,

Boerhaave Haarlem, Cortenbergh Renkum, Zuiderhout Haarlem, Sint Jozef Deventer, Waelwick Ewijk, Irene Dekkerswald Groesbeek, Margriet Nijmegen, Bonnier-Baars Practice Heesch, Laurens Antonius Rotterdam, Rijndam Rehabilitation Rotterdam, Centrum voor Reuma en Revalidatie Rotterdam, Zonnehuis Vlaardingen, Stichting Pieter van Foreest Delft, Florence The Hague, Zonnehuis Amstelveen, Reade Amsterdam, De Volckaert-SBO Oosterhout, Stichting Elisabeth Breda, Thebe Aeneas Breda, De Riethorst-Stromenland Geertruidenberg, Stichting de Bilthuysen Bilthoven, Zorgcombinatie Noorderboog Reggersoord Meppel, Stichting Groenhuysen Roosendaal, Avoord Zorg en Wonen Etten-Leur, Stichting SHDH Janskliniek Haarlem, Stichting Afasietherapie Amsterdam, Rivas Gorinchem, Aafje Rotterdam, De Zellingen Rijkeshove Rotterdam, Saffier de Residentie Mechropa The Hague, Respect Zorggroep Scheveningen The Hague, Revant Rehabilitation Breda, Surplus Zorg Zevenbergen, Careyn Spijkenisse, De Vogellanden Zwolle, De Hoogstraat Utrecht, Woonzorgconcern IJsselheem Zwolle, Osira Amsterdam, Stichting Sint Jacob Jacobkliniek Haarlem, Viattence De Wendhorst Heerde, Zonnehuisgroep IJssel-Vecht Zwolle, Zorgbalans Driehuis, Novicare Best, Libra Zorggroep Blixembosch Eindhoven, Logopedie Zandvoort Zandvoort, Practice M.P. de Boer Haarlem, Brabantzorg Ammerzoden, Zorggroep Elde Bostel, Van Neynseelgroep Den Bosch, Vivent Rosmalen.

Participants and/or their proxies gave written informed consent to participate in the original studies. All SL-therapists and research team members that were involved in the informed consent procedure for these studies were instructed to always include patients and their proxies in the process. Consequently, candidates could discuss their participation with someone they trusted and proxies of the candidate also approved participation. The only exception was when candidates did not have a next of kin. If baseline tests showed a severe aphasia, these patients were not deemed eligible to provide informed consent solely.

In addition to information provided in one or more conversations, candidates were provided with information leaflets containing all details of the studies; one leaflet with all information for the proxies and one simplified "aphasia friendly" version. The informed consent procedure and materials used were approved by the Medical Ethical Committee of Erasmus MC University Medical Center.

By consenting to participate in SPEAK and RATS-3 participants also approved that their anonymized data could be used in post-hoc analyses. Hence, additional consent for this study was not necessary.

Consent for publication

Not applicable.

Competing interests

HEH, EV and MS receive royalties from the ScreeLing, a linguistic screening tool, used as a variable in the prognostic model. The other authors report no disclosures.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Neurology, Erasmus MC University Medical Center, room Nb-324, PO Box 2040, 3000, CA, Rotterdam, the Netherlands. ²Rijndam Rehabilitation, Rotterdam, the Netherlands. ³Department of Public Health, Erasmus MC University Medical Center, Rotterdam, the Netherlands.

⁴Department of Rehabilitation Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands. ⁵Slotervaart Medical Center, Department of Neurology, Amsterdam, the Netherlands.

Received: 12 December 2017 Accepted: 7 October 2018

Published online: 15 October 2018

References

1. Laska AC, Hellblom A, Murray V, Kahan T, Von Arbin M. Aphasia in acute stroke and relation to outcome. *J Intern Med*. 2001;249(11):3505650954–6820;413–22.
2. Engelter ST, Gostynski M, Papa S, Frei M, Born C, Ajdacic-Gross V, Gutzwiller F, Lyrer PA. Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke*. 2006;37(6):1379–84.

3. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014; 35(29):1925–31.
4. Pedersen PM, Vinter K, Olsen TS. Aphasia after stroke: type, severity and prognosis. The Copenhagen aphasia study. *Cerebrovasc Dis*. 2004;17(1): 35–43.
5. Lazar RM, Minzer B, Antonello D, Festa JR, Krakauer JW, Marshall RS. Improvement in aphasia scores after stroke is well predicted by initial severity. *Stroke*. 2010;41(7):1485–8.
6. Maas MB, Lev MH, Ay H, Singhal AB, Greer DM, Smith WS, Harris GJ, Halpern EF, Koroshetz WJ, Furie KL. The prognosis for aphasia in stroke. *J Stroke Cerebrovasc Dis*. 2012;21(5):350–7.
7. Plowman E, Hentz B, Ellis C Jr. Post-stroke aphasia prognosis: a review of patient-related and stroke-related factors. *J Eval Clin Pract*. 2012;18(3):689–94.
8. Flamand-Roze C, Cauquil-Michon C, Roze E, Souillard-Scemama R, Maintigneux L, Ducruex D, Adams D, Denier C. Aphasia in border-zone infarcts has a specific initial pattern and good long-term prognosis. *Eur J Neurol*. 2011;18(12):1397–401.
9. Hoffmann M, Chen R. The spectrum of aphasia subtypes and etiology in subacute stroke. *J Stroke Cerebrovasc Dis*. 2013;22(8):1385–92.
10. Mazaux JM, Lagadec T, de Seze MP, Zongo D, Asselineau J, Douce E, Trias J, Delair MF, Darrigrand B. Communication activity in stroke patients with aphasia. *J Rehabil Med*. 2013;45(4):341–6.
11. de Riesthal M, Wertz R. Prognosis for aphasia: relationship between selected biographical and behavioural variables and outcome and improvement. *Aphasiology*. 2004;18(10):899–915.
12. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
13. El Hachoui H, Lingsma HF, van de Sandt-Koenderman MW, Dippel DW, Koudstaal PJ, Visch-Brink EG. Long-term prognosis of aphasia after stroke. *J Neurol Neurosurg Psychiatry*. 2013;84:310–5.
14. Godecke E, Rai T, Ciccone N, Armstrong E, Granger A, Hankey GJ. Amount of therapy matters in very early aphasia rehabilitation after stroke: a clinical prognostic model. *Semin Speech Lang*. 2013;34(3):129–41.
15. Payabvash S, Kamalian S, Fung S, Wang Y, Passanese J, Kamalian S, Souza LC, Kemmling A, Harris GJ, Halpern EF, et al. Predicting language improvement in acute stroke patients presenting with aphasia: a multivariate logistic model using location-weighted atlas-based analysis of admission CT perfusion scans. *AJNR Am J Neuroradiol*. 2010;31(9):1661–8.
16. Blom-Smink MRMA, de Sandt-Koenderman MWME V, CLJJ K, El Hachoui H, Visch-brink EG, Ribbers GM. Prediction of everyday verbal communicative ability of aphasic stroke patients after inpatient rehabilitation. *Aphasiology*. 2017;31(12):1379–91.
17. Goodglass H, Kaplan E. The assessment of aphasia and related disorders. Philadelphia: Lea and Febiger; 1972.
18. Visch-Brink EG, de Sandt-Koenderman MWE V, El Hachoui H. ScreeLing. Houten: Bohn Stafleu van Loghum; 2010.
19. El Hachoui H, Sandt-Koenderman MW, Dippel DW, Koudstaal PJ, Visch-Brink EG. The ScreeLing: occurrence of linguistic deficits in acute aphasia post-stroke. *J Rehabil Med*. 2012;44(5):429–35.
20. El Hachoui H, Visch-Brink EG, de Lau LM, Van de Sandt-Koenderman MW, Nouwens F, Koudstaal PJ, Dippel DW. Screening tests for aphasia in patients with stroke: a systematic review. *J Neurol*. 2017;264(2):211–20.
21. Nouwens F, Dippel DW, de Jong-Hagelstein M, Visch-Brink EG, Koudstaal PJ, de Lau LM, RATS-investigators. Rotterdam aphasia therapy study (RATS)-3: "the efficacy of intensive cognitive-linguistic therapy in the acute stage of aphasia"; design of a randomised controlled trial. *Trials*. 2013;14:24.
22. Nouwens F, de Lau LML, Visch-Brink EG, Van de Sandt-Koenderman WME, Lingsma HF, Goosen S, Blom DMJ, Koudstaal PJ, Dippel DWJ. Efficacy of early cognitive-linguistic treatment for aphasia due to stroke: a randomised controlled trial (Rotterdam aphasia therapy Study-3). *Eur Stroke J* 2017, 2(2):126–136.
23. Mahoney FI, Barthel DW. Functional evaluation: the BARTHEL index. *Md State Med J*. 1965;14:61–5.
24. Visch-Brink EG, Van de Sandt-Koenderman M, El Hachoui H. ScreeLing. Houten: Bohn Stafleu van Loghum; 2010.
25. Graetz P, De Bleser R, Willmes K. Akense Afasie Test. Nederlandstalige versie. Lisse: Swets & Zeitlinger; 1991.
26. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279–89.
27. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ*. 2009;338:b604.
28. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer; 2009.
29. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.
30. Quinn TJ, Langhorne P, Stott DJ. Barthel index for stroke trials: development, properties, and application. *Stroke*. 2011;42(4):1146–51.
31. El Hachoui H, Lingsma HF, Van de Sandt-Koenderman M, DWJ D, Koudstaal PJ, Visch-Brink EG. Recovery of aphasia after stroke: a 1-year follow-up study. *J Neurol*. 2013;260(1):166–71.
32. van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*. 2000;19(24):3401–15.
33. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P, Group CHDRP. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *Jama*. 2001;286(2):180–7.
34. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56(5):441–7.
35. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*. 2016;74:167–76.
36. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:b375.
37. Pedersen PM, Jorgensen HS, Nakayama H, Raaschou HO, Olsen TS. Aphasia in acute stroke: incidence, determinants, and recovery. *Ann Neurol*. 1995; 38(4):659–66.
38. Prins R, Bastiaanse R. Analysis of spontaneous speech. *Aphasiology*. 2004; 18(12):1075–91.
39. Glize B, Villain M, Richert L, Vellay M, de Gabory I, Mazaux JM, Dehaill P, Sibon I, Laganaro M, Joseph PA. Language features in the acute phase of poststroke severe aphasia could predict the outcome. *Eur J Phys Rehabil Med*. 2017;53(2):249–55.
40. Duffy L, Gajree S, Langhorne P, Stott DJ, Quinn TJ. Reliability (inter-rater agreement) of the Barthel index for assessment of stroke survivors: systematic review and meta-analysis. *Stroke*. 2013;44(2):462–8.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

