Chapter 3

Prognosis of aphasia

Chapter 3.1

Early effect of intra-arterial treatment in ischemic stroke on aphasia recovery in MR CLEAN

Crijnen YS, Nouwens F, de Lau LML, Visch-Brink EG, van de Sandt-Koenderman WME, Berkhemer OA, Fransen PSS, Beumer D, van den Berg LA, Lingsma HF, Roos YBWEM, van der Lugt A, van Oostenbrugge RJ, van Zwam WH, Majoie CBLM, Dippel DWJ.
Neurology. 2016, vol. 86, no. 22, pages: 2049-2055.

ABSTRACT

Objective

To investigate the effect of intra-arterial treatment (IAT) on early recovery from aphasia in acute ischemic stroke. We hypothesized that the early effect of IAT on aphasia is smaller than the effect on motor deficits.

Methods

We included patients with aphasia from the Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands (MR CLEAN), in which 500 patients with a proximal anterior circulation stroke were randomized to usual care plus IAT (<6 hours after stroke, mainly stent retrievers) or usual care alone. We estimated the effect of IAT on the shift on the NIH Stroke Scale (NIHSS) item language and the NIHSS item motor arm at 24 hours and one week after stroke with multivariable ordinal logistic regression as a common odds ratio, adjusted for prognostic variables (acOR). Differences between the effect of IAT on aphasia and on motor deficits were tested in a multilevel model with a multiplicative interaction term.

Results

Of the 288 patients with aphasia, 126 were assigned to IAT and 162 to usual care alone. The acOR for improvement of language score at 24 hours was 1.65 (95% confidence interval, CI: 1.05 to 2.60), and at one week 1.86 (95% CI: 1.18 to 2.94). The acOR for improvement of motor deficit at 24 hours was 2.44 (95% CI: 1.54 to 3.88), and at one week 2.32 (95% CI: 1.43 to 3.77). The effect of IAT on language deficits was significantly different from the effect on motor deficits at 24 hours and one week (p = 0.01 and p = 0.01).

Conclusion

IAT results in better early recovery from aphasia than usual care alone. The early effect of IAT on aphasia is smaller than the effect on motor deficits. This study provides Class II evidence that for patients with acute ischemic stroke IAT increases early recovery from aphasia and that the early effect on aphasia, as measured by the NIHSS, is smaller than the effect on motor deficits.

INTRODUCTION

Recently, several randomized clinical trials showed that intra-arterial treatment (IAT) with retrievable stents for patients with acute ischemic stroke (AIS) caused by a proximal intracranial occlusion in the anterior circulation is safe and improves functional outcome at 90 days.[1-5]

Aphasia is diagnosed in 15% to 40% of patients at ischemic stroke onset.[6-11] Stroke patients with aphasia have increased mortality,[7] decreased functional recovery,[12] and reduced probability to return to work,[13] and they have a higher risk of depression,[8] compared to those without aphasia.[9] Clinical observations and previous studies suggest that language deficits in AIS do not respond as rapidly to reperfusion therapy as other neurologic deficits, especially upper limb paresis.[14-16] To our knowledge, no study thus far has assessed the effect of IAT on aphasia.

Our first aim was to determine whether usual care plus IAT would be more effective than usual care alone for the early recovery from aphasia in patients with AIS. The second aim was to evaluate whether the effect of IAT on early aphasia recovery differed from the effect on early motor deficit recovery.

METHODS

Study design

For this study, we used data from the Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands (MR CLEAN),[1] a randomized, pragmatic phase III, multicenter clinical trial with a Prospective Randomized Open Blinded Endpoint (PROBE) design,[17] among 500 AIS patients with a proximal intracranial arterial occlusion of the anterior circulation. In this trial, usual care plus IAT (mainly with retrievable stents) was compared with usual care alone. Usual care in MR CLEAN could include intravenous administration of alteplase. The MR CLEAN study design has been described in detail elsewhere.[18]

Participants

Patient characteristics and intervention details for MR CLEAN have been extensively described previously.[1] For this study, we selected all patients with aphasia at baseline, affirmed with a deficit on the item language (item 9) of the NIH Stroke Scale (NIHSS). Patients who were comatose at baseline (defined as a score of 3 on the item orientation, item 1A of the NIHSS) were excluded.

Clinical and radiologic assessment

All patients underwent clinical assessment at baseline, which included demographics, risk factors, medical history, pre-stroke modified Rankin Scale score (mRS), and NIHSS score. Assessment of the NIHSS score was repeated after 24 hours and at one week (range 5 to 7 days) or discharge. Investigators were trained to conduct the NIHSS. The NIHSS consists of standardized items, with good interrater reliability.[19, 20] The imaging committee evaluated the baseline vessel images (CT angiography, magnetic resonance angiography, or digital subtraction angiography) to ascertain the location of the occlusion.

Outcomes

The first outcome was the score on the item language of the NIHSS at 24 hours and at one week. This item was scored by asking the patient to name items, perform a complex picture description task (cookie theft), and perform a repetition task, according to the NIHSS manual.[21] Scores on the language item range from 0 to 3, with a score of 0 indicating no aphasia; 1 = mild to moderate aphasia; 2 = severe aphasia; and 3 = mute or global aphasia.

To evaluate the difference with the effect of IAT on motor arm deficits, we used the NIHSS item motor arm at 24 hours and at one week. The item motor arm (item 5) was measured by determining motor arm strength contralateral to the affected hemisphere. The patient was asked to extend the arms 90° and hold this for ten seconds. Scores on this item range from 0 to 4, with a score of 0 indicating no drift; 1 = drift; 2 = some effort against gravity; 3 = no effort against gravity; and 4 = no movement.

Statistical analyses

Missing values for baseline variables that were used to adjust the regression models were imputed with mean or mode, as applicable. Single missing values of items on the NIHSS at 24 hours and at one week were imputed by mode. The percentage of single imputed data was 0.28%. There were no single missing values on the items language and motor arm at 24 hours and at one week. Patients who died within seven days after stroke onset were given the worse score for missing values on the items language (score 3) and motor arm (score 4). Patients who were lost to follow-up because of early discharge and did not die within seven days after stroke onset were not included in the analyses.

All analyses were based on the intention-to-treat principle. The primary effect estimate was the adjusted common odds ratio (acOR) for a shift in the direction of a better outcome on the item language. The acOR estimates the likelihood that IAT would lead to lower NIHSS scores, as compared with usual care alone (shift analysis).[22] This ratio was estimated with multivariable ordinal logistic regression. Estimates were adjusted for age, stroke severity (total NIHSS score) at baseline, history of ischemic stroke, atrial fibrillation, diabetes mellitus, carotid top occlusion, and time from stroke onset to randomization. The acOR for the effect of IAT on the item motor arm was estimated similarly.

In order to summarize the data, we plotted proportions of patients with good outcome at 24 hours and one week. Good outcome was defined, both for aphasia and for motor arm deficit, as a score of 0 or 1 on the NIHSS item. For further analyses, the total distribution of scores was used.

To evaluate whether there is a differential effect of IAT on language versus motor arm recovery, two records per patient were created in the database, one with the language outcome and one with the motor arm outcome. We then fitted a multilevel model with a random intercept for patient, to account for the correlation within patients, and a multiplicative interaction between outcome type and treatment. Since common odds ratios were used, the scales with different ranges could be accurately compared, because they represent weighted averages of odds ratios for each possible dichotomization of the ordinal scale. This analysis was conducted with and without imputed data.

The adjusted and unadjusted odds ratios are reported with 95% confidence intervals (95% CI) to indicate statistical precision. All analyses were performed using Stata/SE statistical package, version 13.1 (StataCorp, College Station, TX).

Standard protocol approvals, registrations, and patient consents

This study is a post-hoc study of MR CLEAN. The trial protocol has been approved by a central medical ethics committee and the research board of each participating center.[18] MR CLEAN is registered in the Dutch trial register (NTR1804) and in the ISRCTN register (ISRCTN10888758). Written informed consent was obtained from all participants or their legal representatives before randomization.

RESULTS

Study population

We identified 289 patients with a language score of >0 at baseline. One comatose patient was excluded. In total, 288 patients were selected for analyses, of whom 126 (44%) had been assigned to the intervention group and 162 (56%) to the control group. One patient received IAT after being assigned to the control group. IAT was never initiated in 12 patients (9.5%) assigned to the intervention group. Baseline characteristics were similar in the two treatment groups (Table 1).

Imputation

After 24 hours and one week, respectively, nine and 37 patients had died and were imputed with a maximum score of 3 on the item language and 4 on the item motor arm. Seven patients at 24 hours and 16 patients at one week were lost to follow-up and were not included in the analyses.

Effect of IAT on language function

There was a shift in the distribution of language scores in favor of IAT at both 24 hours and one week. The acORs for improvement were 1.65 at 24 hours (95% CI: 1.05 to 2.60) and 1.86 at one week (95% CI: 1.18 to 2.94) (Table 2). The shift towards better outcomes in favor of the intervention was consistent for all categories of the item language (Figure 1). At 24 hours, 13% of the patients in the intervention group had no aphasia versus 4% in the control group (acOR: 3.51, 95% CI: 1.28 to 9.67) and at one week these percentages were 21% versus 10% (acOR: 2.45, 95% CI: 1.17 to 5.10) (Table 2).

Table 1. Baseline characteristics of the study population

Characteristics	Intervention (n = 126)	Control (n = 162)
Age in years, median (IQR)	65 (56-76)	66 (58-76)
Male sex, n (%)	75 (60%)	105 (65%)
Total NIHSS score◊, median (IQR)	20 (16-23)	21 (16-23)
NIHSS score item language, n (%)		
1	15 (12%)	13 (8%)
2	41 (33%)	54 (33%)
3	70 (56%)	95 (59%)
NIHSS score item motor arm△, n (%)		
0	3 (2%)	4 (3%)
1	10 (8%)	14 (9%)
2	19 (15%)	14 (9%)
3	14 (11%)	24 (15%)
4	80 (64%)	106 (65%)
Pre-stroke modified Rankin score○, n (%)		
0	104 (83%)	128 (79%)
1	10 (8%)	18 (11%)
2	8 (6%)	9 (6%)
>2	4 (3%)	7 (4%)
History of ischemic stroke, n (%)	13 (10%)	17 (11%)
Atrial fibrillation, n (%)	34 (27%)	51 (25%)
Diabetes mellitus, n (%)	24 (19%)	20 (12%)
Systolic blood pressure, mm Hg, mean (SD)	147 (28)	145 (23)
Location of stroke in left hemisphere, n (%)	114 (91%)	148 (91%)
Carotid top occlusion, n (%)	34 (27%)	43 (27%)
Treatment with IV alteplase, n (%)	112 (89%)	142 (88%)
Time from stroke onset to start of IV alteplase, min		
Median (IQR)	94 (70-108)	100 (65-116)
Time from stroke onset to randomization, min		
Median (IQR)	208 (158-249)	212 (159-264)
Time from stroke onset to IAT, min		
Median (IQR)	266 (215-315)	n.a.

Abbreviations: n = number; IQR = interquartile range; n.a. = not applicable; NIHSS = NIH Stroke Scale; SD = standard deviation; IV = intravenous.

◊ Scores on the NIHSS (a 15-item scale) range from 0 to 42, with higher scores indicating more severe neurologic deficits.

△ NIHSS score on the item motor arm was measured on the arm contralateral to the affected hemisphere.

○ Scores on the modified Rankin scale of functional disability range from 0: no symptoms to 6: death. A score of 2 or less indicates functional independence.

Table 2. Effect of IAT on language function

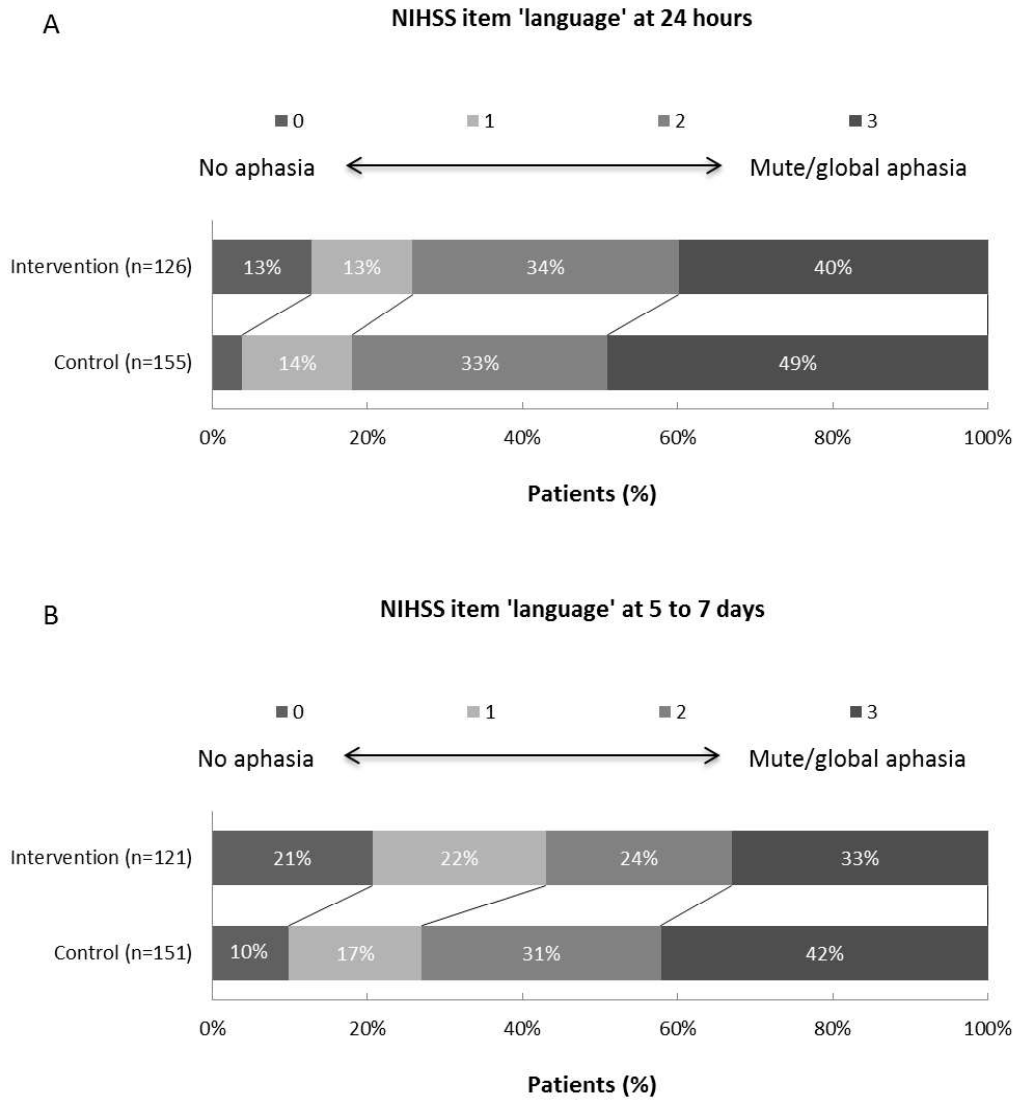
Outcome	Intervention	Control	Effect variable	Unadjusted value [95% CI]	Adjusted value [95% CI]
NIHSS item language at 24 h◊	n = 126	n = 155	Common odds ratio	1.58 [1.02 – 2.44]	1.65 [1.05 – 2.60]
Score of 0, n (%)	16 (13%)	6 (4%)	Odds ratio	3.61 [1.37 – 9.53]	3.51 [1.28 – 9.67]
Score of 0 to 1, n (%)	33 (26%)	28 (18%)	Odds ratio	1.61 [0.91 – 2.85]	1.59 [0.85 – 2.98]
Scores of 0 to 2, n (%)	76 (60%)	79 (51%)	Odds ratio	1.46 [0.91 – 2.35]	1.54 [0.92 – 2.59]
NIHSS item language at 1 week△	n = 121	n = 151	Common odds ratio	1.78 [1.15 – 2.76]	1.86 [1.18 – 2.94]
Score of 0, n (%)	25 (21%)	15 (10%)	Odds ratio	2.36 [1.18 – 4.71]	2.45 [1.17 – 5.10]
Score of 0 to 1, n (%)	52 (43%)	41 (27%)	Odds ratio	2.02 [1.22 – 3.36]	2.21 [1.25 – 3.94]
Scores of 0 to 2, n (%)	81 (67%)	87 (58%)	Odds ratio	1.49 [0.91 – 2.45]	1.55 [0.89 – 2.70]

Abbreviations: IAT = intra-arterial treatment; 95% CI = 95% confidence interval; n = number; NIHSS = NIH Stroke Scale.

◊ The NIHSS score was not available for seven patients who were lost to follow-up at 24 hours and did not die within seven days after stroke.

△ The NIHSS score was not available for 16 patients who were lost to follow-up at one week and did not die within seven days after stroke.

Figure 1. Distribution of scores on the item language of the NIH Stroke Scale (NIHSS) in the intention-to-treat population

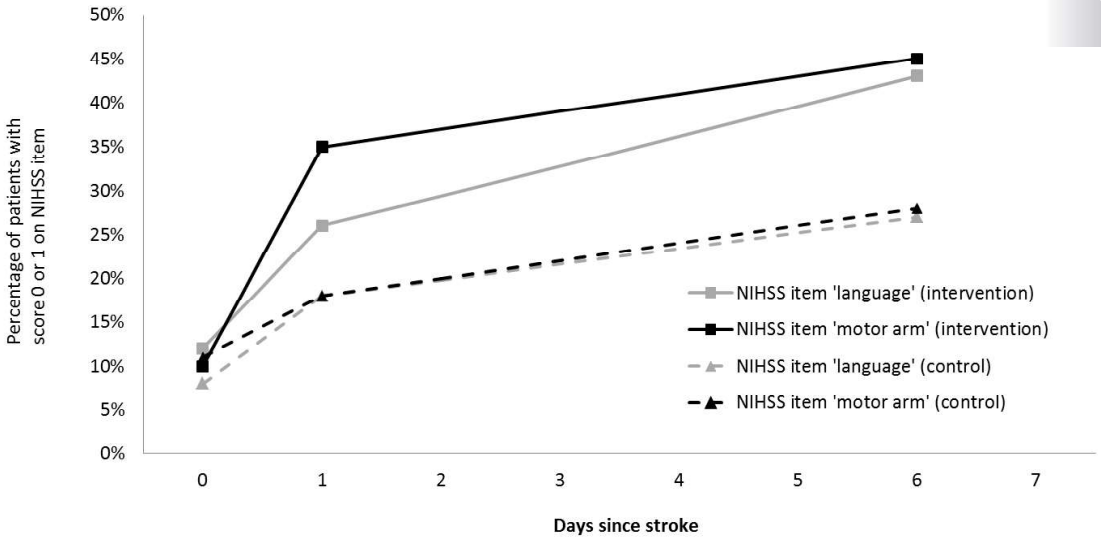


Distribution of scores on the item language of the NIHSS at 24 hours (A) and one week (B) after stroke. Scores range from 0 to 3. At 24 hours, 4% of the patients in the control group had a score of 0.

Effect of IAT on aphasia versus motor arm deficits

We visualized the effects of treatment by plotting proportions of patients with good outcome (score 0 or 1 on the items motor arm and language) at 24 hours and one week. At baseline, the proportions of patients with little or no language or motor deficits in the intervention group were similar. At 24 hours, more patients had a good outcome on the item motor arm than on the item language, but at one week these proportions were again equal (Figure 2).

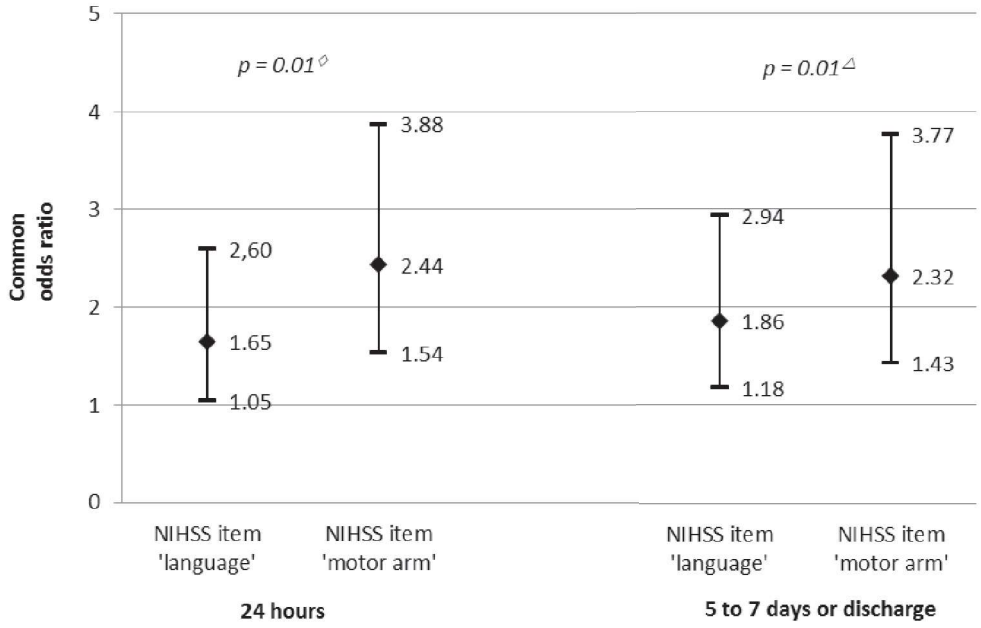
Figure 2. Good outcome scores of language and motor arm



Proportions of patients with good outcome (score 0 or 1) on the item language and on the item motor arm at 24 hours and at one week.
Proportions displayed at day six are based on the NIH Stroke Scale (NIHSS) scores measured at five to seven days or discharge if earlier.

We observed a shift in the distribution of all scores for the NIHSS item motor arm in favor of IAT. Figure 3 shows acORs for the language item and the motor arm item. At 24 hours, the acOR was 1.65 (95% CI: 1.05 to 2.60) for the item language and 2.44 (95% CI: 1.54 to 3.88) for the item motor arm, in favor of the intervention, meaning that chances of improvement of one or more points on the NIHSS are larger for motor function than for aphasia. The difference between these two ratios was statistically significant (p = 0.01). At one week after stroke, the acOR was 1.86 (95% CI: 1.18 to 2.94) for the item language and 2.32 (95% CI: 1.43 to 3.77) for the item motor arm. This difference was also significant (p = 0.01). Treatment effects on the NIHSS item motor leg were similar (data not shown).

Figure 3. Effect of IAT on aphasia versus motor arm deficits



Adjusted common odds ratios for improvement, expressed by a shift on the overall categories of the items language and motor arm at 24 hours and at one week, between intervention and control group.
Abbreviations: NIHSS = NIH Stroke Scale.
◊ *p*-value for interaction between treatment and outcome (language or motor arm) at 24 hours.
△ *p*-value for interaction between treatment and outcome (language or motor arm) at one week.

Repeated analyses without imputed data yielded slightly more significant differential effects (differential effect between scores on items language and motor arm at 24 hours: p = 0.00, and at one week: p = 0.01).

DISCUSSION

This study shows that in patients with AIS caused by a proximal intracranial arterial occlusion of the anterior circulation, greater improvement of aphasia in the first week can be accomplished when IAT is added to usual care within six hours after stroke onset. Second, we showed that early recovery of aphasia was less than the early recovery of motor function.

In the acute phase, rapid recovery can be attained within hours to days by rapid reperfusion of brain tissue.[23] Subacute recovery follows and is thought to be primarily due to neural reorganization, which is complex and can take several weeks to months.[23]

With the addition of IAT to usual care, recovery of aphasia is accelerated. This is important because aphasia can be severely socially disabling and can affect daily life

tremendously.[7-9, 12, 13] Faster recovery will also reduce the substantial costs associated with post-stroke care and communication rehabilitation. The 1-year cost of caring for stroke patients with aphasia was on average $1,700 more than the cost of caring for stroke patients without aphasia.[24]

Visualizing proportions of patients with good outcome on language and motor items after IAT shows that the proportions with good outcome are divergent at 24 hours, but at one week these proportions are similar. However, analyzing the results by ordinal logistic regression, we found that the early effect of IAT on aphasia remains smaller than the effect on motor arm deficit both at 24 hours and at one week, taking into consideration all categories of deficit. Hence, although the outcome at one week is similar for language and motor functioning, the trajectory of recovery differs after IAT from usual care, which is most likely an effect of IAT.

The effect of treatment with intravenous alteplase on early recovery from aphasia compared to other neurologic deficits has been examined previously, but treatment effects are difficult to compare as these studies present results from different time points and had no control group. In a study among 53 patients with an acute middle cerebral artery (MCA) stroke syndrome, aphasia recovered more slowly than limb motor deficit during treatment with tissue plasminogen activator (tPA).[14] In another study in which NIHSS scores were measured 120 minutes after intravenous tPA treatment in 113 patients with MCA occlusion it was found that aphasia responded less than the other impairments.[15] A retrospective cohort study among 243 patients showed a better recovery from other neurologic deficits than aphasia at 24 hours in patients with severe strokes.[16] On the other hand, in another study, similar improvement of aphasia and limb motor deficit was found at 24 hours and one week after stroke in 109 patients who were mainly treated with intravenous thrombolysis.[10] This last study is difficult to compare with the other studies because of the use of composite NIHSS scores by combining the language item with items for cognitive functioning, which were not specifically designed to test language.

There is a tight link between language and motor systems.[25-27] The recovery of these two systems operates on similar principles,[28] so theoretically, aphasia and motor arm deficit contralateral to the affected hemisphere would be expected to show the same recovery pattern. Clinical observations, however, suggest that language deficits in AIS do not respond as rapidly to IAT as motor deficits, which was confirmed by our findings. The most plausible explanation is that the recovery of language processing is more complex than the recovery of the measured motor functions.[23] There is increasing evidence for a neural multifunctionality in the recovery from aphasia, i.e. an interaction between the neural networks engaged in linguistic and nonlinguistic cognitive and emotional functions.[29] Further improvement of language deficits may require time and language therapy. It is reassuring, however, that the proportions of patients with good outcome are similar for motor and language deficits at one week after stroke.

It is remarkable that 58% of the AIS patients in MR CLEAN (288 of 500) had aphasia, compared to 15% to 40% in earlier studies.[6-11] A probable explanation is that only patients with a proven proximal occlusion were included in the present study, while in other studies imaging of intracranial vessels was not routinely performed, resulting in inclusion of patients with more distal occlusions. It is known that the more proximal the occlusion, the higher the risk of aphasia, especially in case of an occlusion of the MCA.[10]

While other studies have reported left lateralized language functioning in at least 96% of the individuals, in the present study only 91% of the aphasic patients had a stroke in the left hemisphere.[23, 30] This implies an uncommonly high proportion of patients with crossed aphasia in our study. However, in the previously cited study among 109 aphasia patients, a similar percentage of patients with aphasia due to left hemisphere stroke of 94% was found.[10] The authors ascribe this to the very early evaluation of their study, as crossed aphasia tends to recover more rapidly.[31-35] In the current study, language deficits were also evaluated in an early stage.

This study has several methodologic limitations. At first, randomization was slightly unbalanced for this post-hoc analysis, because of block size and multiple stratifications in MR CLEAN. This resulted in more patients in the control group than in the intervention group.

Second, in our study the presence of pre-stroke aphasia and pre-stroke motor deficits were not documented. Although we could not rule out preexisting aphasia and motor deficits, the pre-stroke mRS score of 0 in 80% of the patients and the rate of 90% without previous stroke suggests that pre-stroke aphasia and pre-stroke motor deficits were not likely. Higher pre-stroke mRS scores were evenly distributed between the intervention and control group.

Third, the follow-up time of at most seven days after stroke is relatively short to study recovery of neurologic deficits. However, in this first week great improvement of aphasia was observed, especially in patients who were treated with IAT. This therapy induces early reperfusion, occurring within the first hours after stroke. The first week after stroke already gives a good impression of the effect of IAT on the recovery from aphasia. Differences in functioning after this period can be also attributed to adaptation, learning, or rehabilitation, which obscures the effect of IAT on these changes. However, as the recovery from language deficits can be observed up to several weeks after stroke onset, it would have been worthwhile to extend the follow-up period in subsequent studies.[36]

Finally, this study was not specifically designed to investigate and compare language and motor deficits as it is a post-hoc analysis of a randomized trial. The NIHSS provides a coarse categorization for aphasia severity with only four categories, designed to merely detect aphasia and roughly assess the severity. Nevertheless, NIHSS examination is proven to be reliable in the setting of acute stroke evaluation.[19] The assessment of language and motor function is among the most reliable test items.[20] In addition, more advanced tests are difficult to apply in the acute phase after stroke, because these tests are more time-consuming. However, more specific research of treatment effect on different language modalities is needed. A well-known and validated screening tool that can be administered without special training is the Frenchay Aphasia Screening Test, measuring comprehension, expression, reading, and writing in only ten minutes.[37]

Methodologic strengths of this study are the multicenter character, randomized treatment group assignments, and open-label treatment. The broad inclusion criteria led to a wide generalizability of our results. Research on the implementation of IAT is ongoing. Although the positive effect of IAT on functional outcome has been shown, the effects on specific domains are unknown. As yet, studies on aphasia recovery are scarce, so these results are fairly unique.

CONCLUSION

We found that IAT within six hours after stroke onset results in better early recovery from aphasia than usual care alone in patients with a proximal intracranial arterial occlusion of the anterior circulation. Our hypothesis that the very early effect of IAT on aphasia would be smaller than the effect on motor deficit was confirmed, supporting the notion that language, as a more complex function, recovers more slowly than motor function.

REFERENCES

1. Berkhemer OA, Fransen PSS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A Randomized Trial of Intraarterial Treatment for Acute Ischemic Stroke. *New England Journal of Medicine*. 2015;372:11-20
2. Campbell BC, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med*. 2015;372:1009-1018
3. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med*. 2015;372:1019-1030
4. Jovin TG, Chamorro A, Cobo E, de Miquel MA, Molina CA, Rovira A, et al. Thrombectomy within 8 hours after symptom onset in ischemic stroke. *N Engl J Med*. 2015;372:2296-2306
5. Saver JL, Goyal M, Bonafe A, Diener HC, Levy EI, Pereira VM, et al. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. *N Engl J Med*. 2015;372:2285-2295
6. Inatomi Y, Yonehara T, Omiya S, Hashimoto Y, Hirano T, Uchino M. Aphasia during the acute phase in ischemic stroke. *Cerebrovasc Dis*. 2008;25:316-323
7. Laska AC, Hellblom A, Murray V, Kahan T, Von Arbin M. Aphasia in acute stroke and relation to outcome. *J Intern Med*. 2001;249:413-422
8. Kauhanen ML, Korpelainen JT, Hiltunen P, Maatta R, Mononen H, Brusin E, et al. Aphasia, depression, and non-verbal cognitive impairment in ischaemic stroke. *Cerebrovasc Dis*. 2000;10:455-461
9. Engelter ST, Gostynski M, Papa S, Frei M, Born C, Ajdacic-Gross V, et al. Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke*. 2006;37:1379-1384
10. Denier C, Flamand-Roze C, Dib F, Yeung J, Solignac M, Bayon de la Tour L, et al. Aphasia in stroke patients: early outcome following thrombolysis. *Aphasiology*. 2015;29:442-456
11. Croquelois A, Bogousslavsky J. Stroke aphasia: 1,500 consecutive cases. *Cerebrovasc Dis*. 2011;31:392-399
12. Paolucci S, Antonucci G, Pratesi L, Traballesi M, Lubich S, Grasso MG. Functional outcome in stroke inpatient rehabilitation: predicting no, low and high response patients. *Cerebrovasc Dis*. 1998;8:228-234
13. Black-Schaffer RM, Osberg JS. Return to work after stroke: development of a predictive model. *Arch Phys Med Rehabil*. 1990;71:285-290
14. Felberg RA, Okon NJ, El-Mitwalli A, Burgin WS, Grotta JC, Alexandrov AV. Early dramatic recovery during intravenous tissue plasminogen activator infusion: clinical pattern and outcome in acute middle cerebral artery stroke. *Stroke*. 2002;33:1301-1307
15. Mikulik R, Dusek L, Hill MD, Fulep E, Grotta JC, Ribo M, et al. Pattern of response of National Institutes of Health Stroke Scale components to early recanalization in the CLOTBUST trial. *Stroke*. 2010;41:466-470

16. Kremer C, Kappelin J, Perren F. Dissociation of severity of stroke and aphasia recovery early after intravenous recombinant tissue plasminogen activator thrombolysis. *J Clin Neurosci*. 2014;21:1828-1830
17. Hansson L, Hedner T, Dahlof B. Prospective randomized open blinded end-point (PROBE) study. A novel design for intervention trials. Prospective Randomized Open Blinded End-Point. *Blood Press*. 1992;1:113-119
18. Fransen PS, Beumer D, Berkhemer OA, van den Berg LA, Lingsma H, van der Lugt A, et al. MR CLEAN, a multicenter randomized clinical trial of endovascular treatment for acute ischemic stroke in the Netherlands: study protocol for a randomized controlled trial. *Trials*. 2014;15:343
19. Lyden P, Brott T, Tilley B, Welch KM, Mascha EJ, Levine S, et al. Improved reliability of the NIH Stroke Scale using video training. NINDS TPA Stroke Study Group. *Stroke*. 1994;25:2220-2226
20. Goldstein LB, Bertels C, Davis JN. INterrater reliability of the nih stroke scale. *Archives of Neurology*. 1989;46:660-662
21. National Institute of Health, National Institute of Neurological Disorders and Stroke. Stroke Scale.http://www.ninds.nih.gov/doctors/NIH_Stroke_Scale.pdf. Accessed June 10, 2015
22. Saver JL. Novel end point analytic techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials. *Stroke*. 2007;38:3055-3062
23. Marsh EB, Hillis AE. Recovery from aphasia following brain injury: the role of reorganization. *Prog Brain Res*. 2006;157:143-156
24. Ellis C, Simpson AN, Bonilha H, Mauldin PD, Simpson KN. The one-year attributable cost of poststroke aphasia. *Stroke*. 2012;43:1429-1431
25. Gentilucci M, Benuzzi F, Gangitano M, Grimaldi S. Grasp with hand and mouth: a kinematic study on healthy subjects. *J Neurophysiol*. 2001;86:1685-1699
26. Gentilucci M, Dalla Volta R. Spoken language and arm gestures are controlled by the same motor control system. *Q J Exp Psychol (Hove)*. 2008;61:944-957
27. Rizzolatti G, Arbib MA. Language within our grasp. *Trends Neurosci*. 1998;21:188-194
28. Harnish S, Meinzer M, Trinastic J, Fitzgerald D, Page S. Language changes coincide with motor and fMRI changes following upper extremity motor therapy for hemiparesis: a brief report. *Brain Imaging Behav*. 2014;8:370-377
29. Cahana-Amitay D, Albert ML. Neuroscience of aphasia recovery: the concept of neural multifunctionality. *Curr Neurol Neurosci Rep*. 2015;15:41
30. Marien P, Paghera B, De Deyn PP, Vignolo LA. Adult crossed aphasia in dextrals revisited. *Cortex*. 2004;40:41-74
31. Alexander MP, Annett M. Crossed aphasia and related anomalies of cerebral organization: case reports and a genetic hypothesis. *Brain Lang*. 1996;55:213-239
32. Carr MS, Jacobson T, Boller F. Crossed aphasia: analysis of four cases. *Brain Lang*. 1981;14:190-202
33. Croquelois A, Wintermark M, Reichhart M, Meuli R, Bogousslavsky J. Aphasia in hyperacute stroke: language follows brain penumbra dynamics. *Ann Neurol*. 2003;54:321-329
34. Joanette Y, Puel M, Nespoulous JL, Rascol A, Lecours AR. [Crossed aphasia in right-handed patients. I. Review of the literature]
Aphasie croisee chez les droitiers. I. Revue de la litterature. *Rev Neurol (Paris)*. 1982;138:575-586
35. Pedersen PM, Vinter K, Olsen TS. Aphasia after stroke: type, severity and prognosis. The Copenhagen aphasia study. *Cerebrovasc Dis*. 2004;17:35-43
36. El Hachioui H, Lingsma HF, van de Sandt-Koenderman ME, Dippel DW, Koudstaal PJ, Visch-Brink EG. Recovery of aphasia after stroke: a 1-year follow-up study. *J Neurol*. 2013;260:166-171

37. Enderby PM, Wood VA, Wade DT, Hewer RL. The Frenchay Aphasia Screening Test: a short, simple test for aphasia appropriate for non-specialists. *Int Rehabil Med*. 1987;8:166-170

Chapter 3.2

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

Nouwens F, Visch-Brink EG, El Hachioui H, Lingsma HF, van de Sandt-Koenderman WME, Dippel DWJ, Koudstaal PJ, de Lau LML.
Submitted.

ABSTRACT

Objective

To externally validate the SPEAK model for the prediction of long-term outcome of aphasia caused by stroke.

Methods

We used data from 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 (score of 4 or 5) and poor outcome (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 patients (68%) had a good outcome, whereas the model predicted 88%. Discrimination of the model was good with an AUC of 0.87 (95% CI: 0.81 to 0.94), but calibration was unsatisfactory. The model overestimated the probability of good outcome (calibration-in-the-large α = -1.98) and the effect of the predictors was weaker in the validation data than in the derivation data (calibration slope β = 0.88). We therefore recalibrated the model to correctly predict good outcome at six months.

Conclusion

After further external validation, the updated SPEAK model, SPEAK-6, may be used in daily practice to discriminate between stroke patients with good and patients with poor outcome of aphasia at six months after stroke. The original model, renamed SPEAK-12, needs further external validation. This study provides Class II evidence that the SPEAK model has good discriminative properties.

INTRODUCTION

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 recovery. Adequate personal prognosis may also contribute to optimizing individual care, which is important as medical and paramedical care becomes increasingly personalized.[3] Prediction of aphasia outcome in aphasia due to stroke 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 aphasia recovery 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] 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 after 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 often used for rating communicative ability in (semi-) 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 used 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) (Box 1). This model explained 55.7% of the variance in the dataset. Internal validity of the model was good, with an AUC (area under the receiver operation characteristic (ROC) curve) of 0.89.[13]

Box 1. The SPEAK model

$$P\,(\text{ASRS} = 4 \text{ or } 5) = e^{y} / (1 + e^{y})$$

y = 2.04 + 0.27(Phonology score) + 0.10(Barthel score) – 0.06(age) – 0.76(education level) + 0.27(cardio-embolic infarction) + 2.18(intracerebral hemorrhage)
e = 2.718 (constant)

Variables:
Phonology Score: score on ScreeLing subpart *Phonology* (score range: 0-24)
Barthel score: score on *Barthel Index* (score range: 0-20)
Age: age at stroke
Educational level: high = 0 (junior high school or middle vocational education up to university), low = 1 (unfinished elementary school up to sophomore high school or lower vocational education)
Cardio-embolic infarction: yes = 1, no = 0
Intracerebral hemorrhage: yes = 1, no = 0

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.

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	First-ever aphasia due to stroke
	Aphasia ascertained by a speech and language therapist using the 36-item Token Test[23] and a score <5 on the ASRS	Aphasia ascertained by a neurologist and a speech and language therapist
	Testable with the ScreeLing	A score below the cut-off point of the Token Test and/or the ScreeLing
	Within two weeks of stroke onset	Within two to six days of stroke onset
	Age between 18 and 85	Adult
	Language near-native Dutch	Language near-native Dutch
	A life expectancy of >six months	
	Able to tolerate intensive treatment	
Exclusion	A subarachnoid or subdural hemorrhage	
	Success or feasibility of intensive language treatment was severely threatened by:	Presence of one of the following criteria:
	▸ Severe dysarthria	▸ Severe dysarthria
	▸ Premorbid dementia	▸ Pre-stroke dementia (suspected or confirmed)
	▸ Illiteracy	▸ Illiteracy
	▸ Severe developmental dyslexia	▸ Developmental dyslexia
	▸ Severe visual perceptual disorders	▸ Severe perceptual disorders of vision and hearing
	▸ Recent psychiatric history	▸ Psychiatric history

Prognostic variables

Patients with aphasia due to stroke were included in RATS-3 within two weeks of stroke. At inclusion, the following baseline variables were recorded: age, sex, education level, stroke type (cerebral infarction or intracerebral hemorrhage) and 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.[24] All participants were tested with the ScreeLing to detect potential deficits in the basic linguistic components.[19, 25] Spontaneous speech samples were collected with semi-standardized interviews according to the Aachen Aphasia Test manual.[26] 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 one year. In RATS-3 follow-up was at four weeks, three and six months after randomization. ASRS scores from the RATS-3 cohort at six months after randomization were used as outcome in the analysis, as this was closest in time to the original model.

Statistical analyses

Outcome in the RATS-3 cohort was divided in good (ASRS of 4 or 5) or poor (ASRS <4). To validate the SPEAK model we assessed discrimination and calibration.[3, 12, 27-29] For both analyses predicted probability of a good outcome was calculated using the SPEAK model (Box 1).

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. To construct a calibration plot, we ordered the predicted probabilities of good aphasia outcome ascendingly and formed five equally large groups. Per group, the mean probability of a good outcome at six 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 recalibrate the prognostic model by adjusting the intercept.

Handling of missing data

For participants with missing outcome scores at six months, scores at three months after randomization were used. If no scores were available at three 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 six 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 three months was also not available. For 7 participants we used ASRS scores at three months to impute missing values at six 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.

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 in years, mean (SD)	67 (15)	65 (12)
Female sex, n (%)	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 in days, mean (range)	4 (2-6)	8 (1-18)
Barthel Index, median (IQR) ○	15 (8-20)	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

Abbreviations: n = number; SD = standard deviation; IQR = interquartile range; ASRS = Aphasia Severity Rating Scale.

◊ High = senior vocational education, higher education or university.

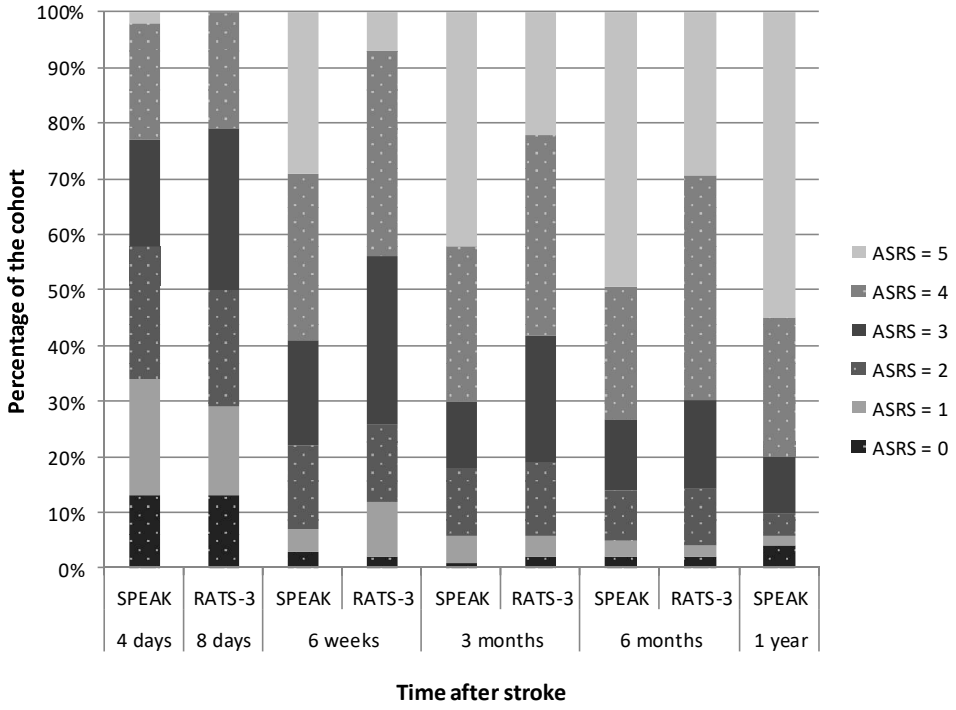
△ Low = no/unfinished elementary school, elementary school, unfinished junior secondary vocational education or junior secondary vocational education.

□ ScreeLing Phonology scores range from 0 to 24.

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

In the derivation SPEAK cohort (n = 130), 11% of the patients had an ASRS score of 4 or 5 at baseline (four days after stroke) and 78% had a good outcome after one year. In the RATS-3 cohort we found a proportion of 21% with a score of 4 or 5 at baseline (eight days after inclusion) and 68% at six 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 Figure 1.

Figure 1. ASRS scores over time in SPEAK and RATS-3

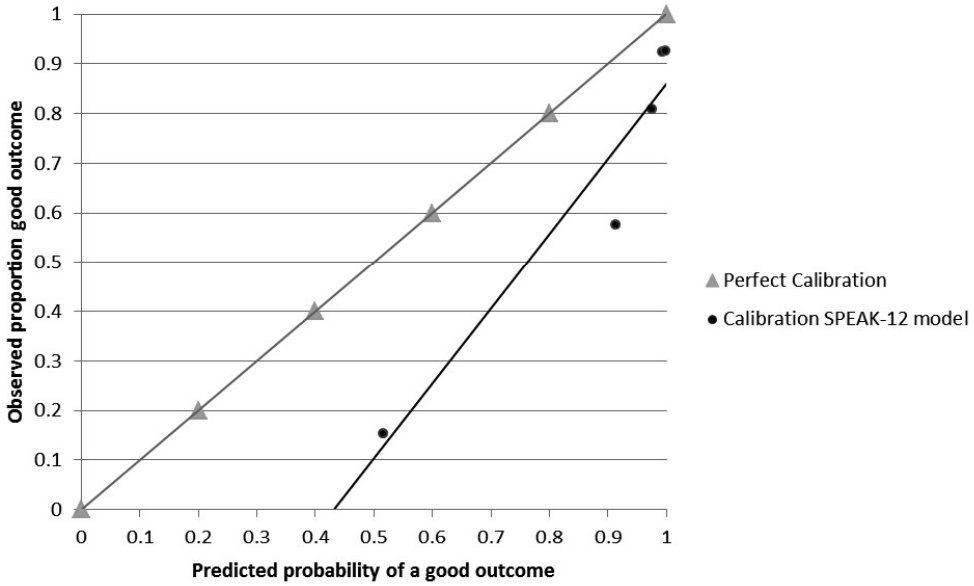


ASRS scores: 5 = minimal discernible speech handicap, some subjective difficulties that are not obvious to the listener; 4 = some obvious loss of fluency in speech or facility of comprehension, without significant limitation in ideas expressed or form of expression; 3 = able to discuss almost all everyday 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.

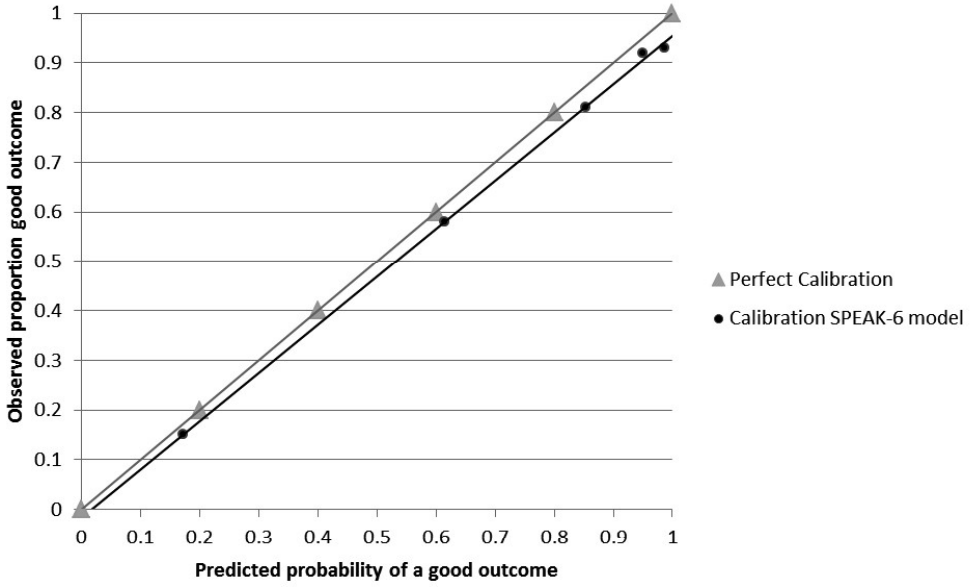
Discrimination of the SPEAK model was good, with an AUC of 0.87 (95% confidence interval, 95% CI: 0.81 to 0.94). In Figure 2A, the grey line depicts calibration of a hypothetically perfect model and the 5 dots represent calibration values in the five subgroups of patients, ordered by increasing predicted probabilities and plotted against the actual proportions of good outcome. The mean predicted probability of good aphasia outcome at one year was 88%, while the observed percentage was 68%, but this was measured at six months. The SPEAK model was too optimistic in predicting good aphasia outcome, with calibration-in-the-large of α = -1.98. The calibration slope of β = 0.88 indicated that the predictor effects were slightly weaker in the validation data than in the derivation data.

Figure 2. Calibration plots of the SPEAK model and updated SPEAK model with predicted probabilities and observed proportions of good aphasia outcome

A. Calibration plot of the original SPEAK model, SPEAK-12



B. Calibration plot of the updated SPEAK model, SPEAK-6



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

Box 2. The updated SPEAK model, SPEAK-6

$$P(\text{ASRS} = 4 \text{ or } 5) = e^{y} / (1 + e^{y})$$

y = 0.06 + 0.27(Phonology score) + 0.10(Barthel score) – 0.06(age) – 0.76(education level) + 0.27(cardio-embolic infarction) + 2.18(intracerebral hemorrhage)
e = 2.718 (constant)

Variables:
Phonology Score: score on ScreeLing subpart *Phonology* (score range: 0-24)
Barthel score: score on *Barthel Index* (score range: 0-20)
Age: age at stroke
Educational level: high = 0 (junior high school or middle vocational education up to university), low = 1 (unfinished elementary school up to sophomore high school or lower vocational education)
Cardio-embolic infarction: yes = 1, no = 0
Intracerebral hemorrhage: yes = 1, no = 0

DISCUSSION

We aimed to externally validate the published SPEAK model for the long-term prognosis of aphasia due to stroke using data from an independent cohort of stroke patients with aphasia, RATS-3. The SPEAK model performed very well in terms of discrimination. 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 one year in SPEAK and six months in RATS-3. Therefore, we proposed an updated version of the SPEAK model for the prediction of outcome at six 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 recovery from aphasia due to 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 therapy (SLT) on aphasia recovery.[14] The authors found that the amount of SLT, added to baseline aphasia severity and baseline stroke disability significantly affected communication four to five weeks after stroke. Baseline variables were recorded within two 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.

In order for a prognostic model to be completely valid and reliable, it is important to evaluate the clinical applicability and generalizability of the model.[30] 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.[31]

Our study is the first to validate a model for the prognosis of aphasia outcome in an independent cohort. Determining 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, 27, 28, 30] 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 four weeks of either intensive cognitive-linguistic treatment or no treatment, starting within two 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 four days after onset and in RATS-3 after eight 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.[32] Hence, these predictors might have different effects in the RATS-3 cohort, as represented in the suboptimal calibration slope.

Third and most importantly, calibration may likely have been influenced by a different follow-up duration, which was six months in RATS-3 versus one year in SPEAK. In SPEAK, ASRS scores improved significantly up to six months after aphasia onset, but no significant improvement was found between six and twelve months.[32] 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 on the ASRS was found between six and twelve months after stroke, some improvement still occurred.[32] Of the participants from SPEAK 74% had an ASRS score of 4 or 5 at six 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 twelve months in the RATS-3 cohort, because of the small, but apparent recovery between six and twelve months after stroke. We therefore suggest an

updated version, SPEAK-6, to predict outcome at six months. More extensive updating could imply refitting the models to the new dataset, to obtain new model coefficients. [33-35] However, as the model discrimination was good, we updated only the intercept to make the model applicable to predict outcome at six months, when the average probability of a good outcome is lower than at one 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.[36]

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.[27] Therefore, as in all clinical trials, one must be careful in generalizing the results to all stroke patients with aphasia.[37]

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.[29, 37] 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.

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, translation into 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]

Finally, the RATS-3 database contained several missing values. Of the participants who refused evaluation at six 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%).[28] 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 recovery after one 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 six 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 recovery at six months after stroke.

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;249:413-422
2. Engelter ST, Gostynski M, Papa S, Frei M, Born C, Ajdacic-Gross V, et al. Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke*. 2006;37:1379-1384
3. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35:1925-1931
4. Pedersen PM, Vinter K, Olsen TS. Aphasia after stroke: type, severity and prognosis. The Copenhagen aphasia study. *Cerebrovasc Dis*. 2004;17:35-43
5. Lazar RM, Minzer B, Antoniello D, Festa JR, Krakauer JW, Marshall RS. Improvement in aphasia scores after stroke is well predicted by initial severity. *Stroke*. 2010;41:1485-1488
6. Maas MB, Lev MH, Ay H, Singhal AB, Greer DM, Smith WS, et al. The Prognosis for Aphasia in Stroke. *J Stroke Cerebrovasc Dis*. 2010
7. Plowman E, Hentz B, Ellis C. Post-stroke aphasia prognosis: a review of patient-related and stroke-related factors. *J Eval Clin Pract*. 2011
8. Flamand-Roze C, Cauquil-Michon C, Roze E, Souillard-Scemama R, Maintigneux L, Ducreux D, et al. Aphasia in border-zone infarcts has a specific initial pattern and good long-term prognosis. *Eur J Neurol*. 2011;18:1397-1401
9. Hoffmann M, Chen R. The spectrum of aphasia subtypes and etiology in subacute stroke. *J Stroke Cerebrovasc Dis*. 2013;22:1385-1392
10. Mazaux JM, Lagadec T, de Seze MP, Zongo D, Asselineau J, Douce E, et al. Communication activity in stroke patients with aphasia. *J Rehabil Med*. 2013;45:341-346
11. de Riesthal M, Wertz R. Prognosis for aphasia: Relationship between selected biographical and behavioural variables and outcome and improvement. *Aphasiology*. 2004;18:899-915
12. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605
13. El Hachioui 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;84:310-315
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;34:129-141
15. Payabvash S, Kamalian S, Fung S, Wang Y, Passanese J, Kamalian S, 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;31:1661-1668
16. Blom-Smink MRMA, van de Sandt-Koenderman MWME, Kruitwagen CLJJ, El Hachioui H, Visch-Brink EG, Ribbers GM. Prediction of everyday verbal communicative ability of aphasic stroke patients after inpatient rehabilitation. *Aphasiology*. 2017:1-13
17. Goodglass H, Kaplan E. *The assessment of aphasia and related disorders.* . Philadelphia: Lea and Febiger; 1972.
18. Visch-Brink EG, Van de Sandt-Koenderman MWE, El Hachioui H. *ScreeLing*. Houten: Bohn Stafleu van Loghum; 2010.
19. El Hachioui 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;44:429-435
20. El Hachioui H, Visch-Brink EG, de Lau LM, van de Sandt-Koenderman MW, Nouwens F, Koudstaal PJ, et al. Screening tests for aphasia in patients with stroke: a systematic review. *J Neurol*. 2016
21. Nouwens F, Dippel DW, de Jong-Hagelstein M, Visch-Brink EG, Koudstaal PJ, de Lau LM, et al. 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: 24:1-8
22. Nouwens F, de Lau LML, Visch-Brink EG, van de Sandt-Koenderman WME, Lingsma HF, Goosen S, et al. Efficacy of early cognitive-linguistic treatment for aphasia due to stroke: A randomised controlled trial (Rotterdam Aphasia Therapy Study-3). *European Stroke Journal*. 2017:2396987317698327
23. De Renzi E, Faglioni P. Normative data and screening power of a shortened version of the Token Test. *Cortex*. 1978;14:41-49
24. Mahoney FI, Barthel DW. Functional evaluation: The Barthel Index. *Md State Med J*. 1965;14:61-65
25. Visch-Brink EG, Van de Sandt-Koenderman M, El Hachioui H. *ScreeLing*. Houten: Bohn Stafleu van Loghum; 2010.
26. Graetz P, De Bleser R, Willmes K. *Aachen Aphasia Test, Dutch version [Akense Afasie Test. Nederlandstalige versie].* Lisse: Swets & Zeitlinger; 1991.
27. 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;68:279-289
28. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338:b604
29. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.* New York: Springer; 2009.
30. 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
31. Quinn TJ, Langhorne P, Stott DJ. Barthel Index for Stroke Trials: Development, Properties, and Application. *Stroke*. 2011;42:1146-1151
32. El Hachioui H, Lingsma HF, van de Sandt-Koenderman M, Dippel DWJ, Koudstaal PJ, Visch-Brink EG. Recovery of aphasia after stroke: a 1-year follow-up study. *J Neurol*. 2013;260:166-171
33. van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*. 2000;19:3401-3415
34. 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;286:180-187

35. 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;56:441-447
36. 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;74:167-176
37. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:b375
38. Prins R, Bastiaanse R. Review. *Aphasiology*. 2004;18:1075-1091