An MRI Rating Scale for Amyloid-Related Imaging Abnormalities with Edema or Effusion

F. Barkhof, M. Daams, P. Scheltens, H.R. Brashear, H.M. Arrighi, A. Bechten, K. Morris, M. McGovern and M.P. Wattjes

*AJNR Am J Neuroradiol* 2013, 34 (8) 1550-1555
doi: https://doi.org/10.3174/ajnr.A3475
http://www.ajnr.org/content/34/8/1550
ABSTRACT

BACKGROUND AND PURPOSE: Immune therapy against amyloid-β appears to be a promising target in Alzheimer disease. However, a dose-related risk for ARIA on FLAIR images thought to represent parenchymal vasogenic edema or sulcal effusion (termed “ARIA-E”), has been observed in clinical trials. To assess the intensity of ARIA-E presentation, an MR imaging scale that is both reproducible and easily implemented would assist in monitoring and evaluating this adverse event.

MATERIALS AND METHODS: On the basis of a review of existing cases from a phase II bapineuzumab study, a scale was constructed with a 6-point score for the 6 regions on each side of the brain (range, 0—60). Scores would be obtained for both parenchymal and sulcal hyperintensities and frequently co-occurring gyral swelling. Inter-rater reliability between 2 neuroradiologists was evaluated in 20 patients, 10 with known ARIA-E and 10 without, by using the intraclass correlation coefficient.

RESULTS: The 2 raters had excellent agreement in the identification of ARIA-E cases. A high inter-rater agreement was observed for scores of parenchymal hyperintensity (ICC 0.83; 95% CI, 48—96) and sulcal hyperintensity (ICC 0.89; 95% CI, 63—97) and for the combined scores of the 2 ARIA-E findings (ICC 0.89; 95% CI, 62—97). Gyral swelling scores were observed to have lower inter-rater agreement (ICC 0.54; 95% CI, 0.06—0.86).

CONCLUSIONS: The proposed rating scale provides a reliable and easily implemented instrument to grade ARIA-E imaging findings. We currently do not recommend including swelling.

ABBREVIATIONS: Aβ = amyloid-beta; AD = Alzheimer disease; ApoE = apolipoprotein E; ARIA = amyloid-related imaging abnormalities; ICC = intraclass correlation coefficient

Alzheimer disease is a progressive neurodegenerative disease associated with dementia and is histopathologically characterized by cerebral neuronal loss, deposits of extracellular plaques of Aβ, and the intraneural accumulation of hyperphosphorylated τ neurofibrillary tangles.1,2 Treatment strategies targeted against these insults are being investigated; however, to date, no curative treatment exists. Therapies targeting the Aβ plaques have the longest research history, with the first animal models of immunotherapy for AD introduced >10 years ago.3 Several human in vivo trials have been completed or are ongoing using both active and passive immunization strategies for Aβ.4-6 Immunization against Aβ is hypothesized to lead to an immune-mediated cleavage and removal of Aβ depositions in the brain.7 Animal and human in vivo amyloid PET studies have shown that immunization therapy is effective in terms of Aβ removal, and several studies based on active immunization with the full-length Aβ42 peptide suggested clinical benefits.3,8,9

In addition to Aβ removal, MR imaging findings have been observed that are considered likely related to the clearance mechanism.5,6,10 Dose-related findings include vasogenic edema, sulcal effusion, superficial siderosis, and cerebral microbleeds. The latter are also naturally observed in AD, because lobar microbleeds are related to cerebral amyloid angiopathy and AD pathology.5,10,11-15 Because both findings are considered related to amyloid pathology, the term “amyloid-related imaging abnormalities” has been proposed. ARIA is further subdivided into ARIA-H, representing hemosiderin deposits or superficial hemosiderosis, and ARIA-E, representing parenchymal vasogenic edema or sulcal effusion. ARIA-E can present with different imaging features,
such as gyral swelling and sulcal hyperintensity, along with white matter hyperintensity.16

Scoring guidelines and rating scales for the detection of microbleeds have been established and are widely used in research studies.15,17 Given the number of clinical trials in patients with AD targeting Aβ, a standardized assessment of this rather new imaging finding of ARIA-E would be useful to improve our understanding of its risk factors and outcomes. The aim of our study, therefore, was to establish a reproducible, clinically applicable, visual MR imaging rating scale for ARIA-E and to examine its internal validity in terms of inter-rater reliability.

**MATERIALS AND METHODS**

**Patient Population**

All patients included in this study were part of a phase II, multicenter, randomized, double-blind, placebo-controlled multiple ascending dose immunization study by using bapineuzumab, a humanized monoclonal antibody against Aβ.5 The study was conducted at 30 sites in the United States between April 2005 and March 2008. Two hundred thirty-four patients were randomly assigned to receive either intravenous bapineuzumab or a placebo, in a ratio of 8:7, in 1 of 4 sequential dose cohorts (0.15, 0.5, 1.0, or 2.0 mg/kg). The patients had a mean age of 69 years, with slightly more women (55%), predominantly white (96%), often carrying at least 1 allele (65%) and had a mean Mini-Mental State Examination score of 21 at enrollment (Table 1). Four of the 10 patients with ARIA-E were symptomatic on the basis of the Examination score of 21 at enrollment (Table 1).

A summary score can be derived, and we considered 2 approaches. One approach was to sum the regional scores on each side of the brain (L/R) for each type of abnormality, parenchymal or sulcal, with only the highest score of the 2 characteristics contributing to the score of the region. A second score was obtained by summing the highest scores among the 3 characteristics of the region (parenchymal hyperintensity, sulcal hyperintensity, or gyral swelling).

**Inter-Rater Reliability Testing**

To test inter-rater-reliability, 20 pairs of MR imaging scans (baseline and follow-up) of previously identified ARIA-E cases. All cases were preidentified by a central review and comprise a subsample of the previously published ARIA-E studies by Salloway et al5 and Sperling et al.10 ARIA-E was defined according to the guidelines of the Alzheimer’s Association Research Round Table Workgroup16 and included the occurrence of either sulcal or parenchymal hyperintensities. Two experienced neuroradiologists (F.B. and M.P.W.) reviewed the axial FLAIR images for number and extent of the 2 abnormalities in the 6 regions on both sides (left/right [L/R]) as used for the Age-Related White Matter Changes rating scale,18 and they discussed their findings to define a scale based on the number and maximum in-plane cross-sectional diameter of each abnormality in each region. A third characteristic, gyral swelling, was included in the rating scale by using a similar region-based approach.

**RESULTS**

**ARIA-E Scale Description**

The developed rating scale for ARIA-E included both the location and magnitude of presentation of parenchymal hyperintensities, T2*- and T1-weighted images). Regarding the parenchymal abnormalities, high signal on the T2-weighted images could be confused with partial volume averaging effects from the adjacent CSF and was not considered helpful in addition to FLAIR.

**ARIA-E Rating Scale Development**

For the construction of the scale, we used pairs of MR imaging scans (baseline and follow-up) of previously identified ARIA-E cases. All cases were preidentified by a central review and comprise a subsample of the previously published ARIA-E studies by Salloway et al5 and Sperling et al.10 ARIA-E was defined according to the guidelines of the Alzheimer’s Association Research Round Table Workgroup16 and included the occurrence of either sulcal or parenchymal hyperintensities. Two experienced neuroradiologists (F.B. and M.P.W.) reviewed the axial FLAIR images for number and extent of the 2 abnormalities in the 6 regions on both sides (left/right [L/R]) as used for the Age-Related White Matter Changes rating scale,18 and they discussed their findings to define a scale based on the number and maximum in-plane cross-sectional diameter of each abnormality in each region. A third characteristic, gyral swelling, was included in the rating scale by using a similar region-based approach.

A summary score can be derived, and we considered 2 approaches. One approach was to sum the regional scores on each side of the brain (L/R) for each type of abnormality, parenchymal or sulcal, with only the highest score of the 2 characteristics contributing to the score of the region. A second score was obtained by summing the highest scores among the 3 characteristics of the region (parenchymal hyperintensity, sulcal hyperintensity, or gyral swelling).

**Table 1: Summary of baseline information**

| ARIA-E Cases | Non-ARIA Cases | Overall |
|--------------|---------------|---------|
| No. of subjects | 10 | 10 | 20 |
| Age (mean) (± SD) | 69.4 (9.5) | 69.2 (8.4) | 69.3 (8.7) |
| MMSE (mean) (± SD) | 20.2 (3.9) | 21.2 (2.7) | 20.7 (3.3) |
| ADAS-cog (mean) (± SD) | 22.0 (4.3) | 21.8 (5.6) | 21.9 (6.2) |
| DAD (mean) (± SD) | 81.9 (18.8) | 85.2 (11.2) | 83.5 (15.2) |
| Female (No.) (%) | 7 (70.0%) | 4 (40.0%) | 5 (50.0%) |
| ApoE ε4 status | | | |
| ApoE ε4 noncarrier (No.) (%) | 3 (30.0%) | 1 (10.0%) | 4 (20.0%) |
| ApoE ε4 (No.) (%) | 3 (30.0%) | 5 (50.0%) | 8 (40.0%) |
| ApoE ε4 homozygote (No.) (%) | 4 (40.0%) | 4 (40.0%) | 8 (40.0%) |
| Bapineuzumab | | | |
| 0.15 mg/kg (No.) (%) | 1 (10.0%) | 1 (10.0%) | 2 (10.0%) |
| 0.5 mg/kg (No.) (%) | 1 (10.0%) | 2 (20.0%) | 3 (15.0%) |
| 1.0 mg/kg (No.) (%) | 2 (20.0%) | 3 (30.0%) | 5 (25.0%) |
| 2.0 mg/kg (No.) (%) | 6 (60.0%) | 4 (40.0%) | 10 (50.0%) |

**Note:** DAD indicates Disability Assessment for Dementia; ADAS-cog, Alzheimer’s Disease Assessment Scale-cognition (II items); MMSE, Mini-Mental State Examination.
sulcal hyperintensities, and gyral swelling. If ≥1 of those 3 findings was present, the changes were scored according to the anatomic location in terms of lobe and side, resulting in scores for 6 regions bilaterally: frontal lobe, parietal lobe, temporal lobe, occipital lobe, central region (basal ganglia, thalamus, internal and external capsules, corpus callosum, insula), and infratentorial region (brain stem and cerebellum). Within each region, a score of 0–5 was given on the basis of the spatial extent and multifocality of the abnormality. When a finding covered multiple lobes, the maximum in-plane diameter of the abnormality involving that particular lobe was measured and scored accordingly. Figures 1–3 provide examples of assessing the size and extent of the pathologic changes. A total score can be derived by summing up the 12 regional scores (range, 0–60) from the characteristic, with the maximum score defining the regional score. The scoring scheme is summarized in Table 2. The reading procedure for each patient typically took <5 minutes.

**Descriptive Findings**

The 10 patients with AD with ARIA-E were from a phase II study of bapineuzumab with additional descriptions of the study design and results in Salloway et al (2009)5 and the identification of the ARIA-E findings in Sperling et al (2012).10 Seven of the ARIA-E cases were detected on routine MR imaging, 2 were detected through retrospective MR imaging, and 1 was detected on an unscheduled MR imaging due to symptoms. All patients (69.4 ± 9.5 years of age; 7 women and 3 men) who exhibited ARIA-E were diagnosed with probable AD, with a Mini-Mental State Examination score between 16 and 26 at the initial study enrollment and were assigned to the 4 dose cohorts, with 1 patient at 0.15 mg/kg, 1 at 0.5 mg/kg, 2 at 1.0 mg/kg, and 6 at 2.0 mg/kg. The 10 patients with AD without ARIA-E were selected by using a simple random-sample process from among all patients who were not identified with ARIA-E. Those patients without ARIA-E met the same eligibility requirements for AD as those with probable ARIA-E (AD and a Mini-Mental State Examination score between 16 and 26 at enrollment). They had a nearly identical age at enrollment of 69.2 ± 8.4 years, with 4 women and 6 men, and they were assigned to all 4 dose cohorts, with 1 patient at 0.15 mg/kg, 2 at 0.5 mg/kg, 3 at 1.0 mg/kg, and 4 at 2.0 mg/kg.

As shown in Fig 4, the cases used in this study represented a wide range of ARIA-E pathology and illustrate the dynamics of the scale. Among the 5 cases with the highest scores, the score was strongly driven by parenchymal hyperintensity in cases 1 and 2 (with some additional sulcal hyperintensity), whereas sulcal hyperintensity was the major determinant in cases 3, 5, and 10 (with barely any parenchymal hyperintensity in the latter 2). Scores for swelling followed those of sulcal hyperintensity rather than those of parenchymal hyperintensity. Raters provided identical scores.
In this limited sample, there was excellent consensus on the presence or absence of ARIA-E between the raters. Among the 10 ARIA-E cases, the ICC scores for the component characteristics and total scores are reported in Table 3, along with 95% confidence intervals. Excellent inter-rater agreement was found for the 2 elements of ARIA-E, parenchymal hyperintensity (ICC = 0.83) and sulcal hyperintensity (ICC = 0.89) and for a total score based on these 2 findings (ICC = 0.89). Ratings of swelling exhibited only moderate concordance between the raters (ICC = 0.54), with a wider confidence interval, resulting in a slightly lower ICC for the combination of the 3 features (ICC = 0.78).

**Inter-Rater Agreement**

In this limited sample, there was excellent consensus on the presence or absence of ARIA-E between the raters. Among the 10 ARIA-E cases, the ICC scores for the component characteristics and total scores are reported in Table 3, along with 95% confidence intervals. Excellent inter-rater agreement was found for the 2 elements of ARIA-E, parenchymal hyperintensity (ICC = 0.83) and sulcal hyperintensity (ICC = 0.89) and for a total score based on these 2 findings (ICC = 0.89). Ratings of swelling exhibited only moderate concordance between the raters (ICC = 0.54), with a wider confidence interval, resulting in a slightly lower ICC for the combination of the 3 features (ICC = 0.78).

**DISCUSSION**

We present an easily applicable visual rating scale, which allows characterization of possible ARIA-E findings along 6 anatomic regions per hemisphere by using axial FLAIR images. This rating scale did not require extensive training, and both raters considered its application straightforward and easy to use. Of note, we observed a high degree of inter-rater agreement (ICC > 0.8) for the combined ARIA-E findings of sulcal and parenchymal hyperintensities, demonstrating reasonable internal validity. In the limited sample examined, the scale seems to exhibit a good dynamic range, which would allow the distinction between findings that are extensive and findings localized to a small area in a single region.

Gyral swelling was the characteristic with the least agreement between the raters. When swelling was included in the calculation of the total score, the ICC was 0.78, slightly lower when the characteristic was not included (ICC = 0.89). However, the variance in the estimate substantially increased when swelling was included compared with when it was not included. The lower level of agreement for swelling in isolation is not an unexpected finding because the lack of hyperintensity makes the identification of the boundaries of swelling difficult to determine and likely enhances differences between raters. For this reason and because of the difficulty in terms of scoring particularly subtle changes, swelling was not included in the final grading system. More fundamentally, swelling may be truly based on volume increase of the parenchyma, perhaps resulting from a low level of vasogenic edema, or swelling may be due to a low level of sulcal hyperintensity, resulting in effacement of the normal sulcal low CSF signal intensity on FLAIR. These differing etiologies both contribute to the challenge in determining the boundaries and extent of the swelling (Fig 5).

The separate assessment of different ARIA-E manifestations is complicated because all 3 characteristics (parenchymal hyperintensity, sulcal hyperintensity, and gyral swelling) may be present in a given patient, may partly be indistinguishable from each other in terms of MR imaging appearance, and may be observed in the same region; these complications make it difficult to separate the characteristics in terms of spatial extent. For example, a clear differentiation between sulcal and parenchymal hyperintensities may be a challenge. In addition, sulcal hyperintensities and swelling were observed to frequently be associated with each other, and the presence of one may increase the propensity to identify the other. Additionally, they may share the same underlying etiology, as previously discussed. In our opinion, the extent of brain tissue involved is the most important characteristic to classify; thus, a total score might be most relevant, as opposed to the individual component scores (ie, parenchymal hyperintensity, sulcal hyperintensity, or gyral swelling).

A limitation of our study is that our rating scale has only been validated in terms of inter-rater agreement by 2 specialized and
Our proposed visual rating scale for ARIA-E represents an easy-to-apply rating scale with robust characteristics in terms of inter-rater agreement. On the basis of our results, this rating scale could serve as an important and useful instrument to apply in a clinical setting to monitor imaging abnormalities associated with amyloid-lowering treatment. Given the complexity of the MR imaging characteristics and the pathophysiology of ARIA-E, further validation of this visual rating scale is needed by using larger patient groups and including raters with different degrees of expertise. Most important, the scale awaits external validation of its appropriateness to assist in determining the clinical relevance of ARIA-E findings and possible implications for dose adjustment during amyloid-lowering therapy.

ACKNOWLEDGMENTS

The development and validation of an ARIA-E rating scale by Dr Barkhof was sponsored by Janssen Alzheimer Immunotherapy Research & Development LLC and Pfizer Inc. The authors wish to thank Janssen Alzheimer Immunotherapy Research & Development and Pfizer Inc for providing the MR imaging scans, which were collected as part of a phase II study of bapineuzumab IV, and Dr Dirk Knol (Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, the Netherlands) for statistical advice.

REFERENCES

1. McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer’s disease: a report of the NINCDS-ADRDA work group under the auspices of the Department of Health and Human Services Task Force on Alzheimer’s Disease. Neurology 1984;34:939—44
2. Selkoe DJ, Schenk D. Alzheimer’s disease: molecular understanding predicts amyloid-based therapeutics. *Annu Rev Pharmacol Toxicol* 2003;43:545–84
3. Schenk D, Barbour R, Dunn W, et al. Immunization with amyloid-beta attenuated Alzheimer-disease-like pathology in the PDAPP mouse. *Nature* 1999;400:173–77
4. Gilman S, Koller M, Black RS, et al. Clinical effects of Aβ immunization (AN1792) in patients with AD in an interrupted trial. *Neurology* 2005;64:1533–62
5. Salloway S, Sperling R, Gilman S, et al. A phase 2 multiple ascending dose trial of bapineuzumab in mild to moderate Alzheimer disease. *Neurology* 2009;73:2061–70
6. Ostrowitzki S, Deptula D, Thurfjell L, et al. Mechanism of amyloid removal in patients with Alzheimer disease treated with gantenerumab. *Arch Neurol* 2012;69:198–207
7. Weiner HL, Frenkel D. Immunology and immunotherapy of Alzheimer’s disease. *Nat Rev Immunol* 2006;6:404–16
8. Rinne JO, Brooks DJ, Rossor MN, et al. 11C-PIB assessment of change in fibrillar amyloid-beta load in patients with Alzheimer’s disease treated with bapineuzumab: a phase 2, double-blind, placebo-controlled, ascending dose study. *Lancet Neurol* 2010;9:363–72
9. Bayer AJ, Bullock R, Jones RW, et al. Evaluation of the safety and immunogenicity of synthetic Abeta42 (AN1792) in patients with AD. *Neurology* 2005;64:94–101
10. Sperling R, Salloway S, Brooks DJ, et al. Amyloid-related imaging abnormalities in patients with Alzheimer’s disease treated with bapineuzumab: a retrospective analysis. *Lancet Neurol* 2012;11:241–49
11. Nicoll JA, Wilkinson D, Holmes C, et al. Neuropathology of human Alzheimer disease after immunization with amyloid-beta peptide: a case report. *Nat Med* 2003;9:448–52
12. Boche D, Zotova E, Weller RO, et al. Consequence of Abeta immunization on the vasculature of human Alzheimer’s disease brain. *Brain* 2008;131:3299–310
13. Black RS, Sperling RA, Safirstein B, et al. A single ascending dose study of bapineuzumab in patients with Alzheimer disease. *Alzheimer Dis Assoc Disord* 2010;24:198–203
14. Cordonnier C, van der Flier WM. Brain microbleeds and Alzheimer’s disease: innocent observation or key player? *Brain* 2011;134:335–44
15. Greenberg SM, Vernooij MW, Cordonnier C, et al. Cerebral microbleeds: a guide to detection and interpretation. *Lancet Neurol* 2009;8:165–74
16. Sperling RA, Jack CR, Black SE, et al. Amyloid-related abnormalities in amyloid-modifying therapeutic trials: recommendations from the Alzheimer’s Association Research Round Table Workgroup. *Alzheimers Dement* 2011;7:367–85
17. Gregoire SM, Chaudhary UJ, Brown MM, et al. The Microbleed Anatomical Rating Scale (MARS): reliability of a tool to map brain microbleeds. *Neurology* 2009;73:1759–66
18. Wahlund LO, Barkhof F, Fazekas F, et al. A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke* 2001;32:1318–22
19. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996;1:30–46
20. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychological Reports* 1966;19:3–11