



Providing Choice & Value

Generic CT and MRI Contrast Agents



**FRESENIUS
KABI**

CONTACT REP

AJNR

This information is current as
of July 25, 2025.

Automated Volumetric Software in Dementia: Help or Hindrance to the Neuroradiologist?

Jody Tanabe, Maili F. Lim, Siddhant Dash, Jack Pattee,
Brandon Steach, Peter Pressman, Brianne M. Bettcher, Justin
M. Honce, Valeria A. Potigailo, William Colantoni, David
Zander and Ashesh A. Thaker

AJNR Am J Neuroradiol 2024, 45 (11) 1737-1744

doi: <https://doi.org/10.3174/ajnr.A8406>

<http://www.ajnr.org/content/45/11/1737>

Automated Volumetric Software in Dementia: Help or Hindrance to the Neuroradiologist?

 Jody Tanabe,  Maili F. Lim, Siddhant Dash,  Jack Pattee,  Brandon Steach, Peter Pressman, Brianne M. Bettcher,  Justin M. Honce,  Valeria A. Potigailo, William Colantoni,  David Zander, and  Ashesh A. Thaker



ABSTRACT

BACKGROUND AND PURPOSE: Brain atrophy occurs in the late stage of dementia, yet structural MRI is widely used in the work-up. Atrophy patterns can suggest a diagnosis of Alzheimer disease (AD) or frontotemporal dementia (FTD) but are difficult to assess visually. We hypothesized that the availability of a quantitative volumetric brain MRI report would increase neuroradiologists' accuracy in diagnosing AD, FTD, or healthy controls compared with visual assessment.

MATERIALS AND METHODS: Twenty-two patients with AD, 17 with FTD, and 21 cognitively healthy patients were identified from the electronic health systems record and a behavioral neurology clinic. Four neuroradiologists evaluated T1-weighted anatomic MRI studies with and without a volumetric report. Outcome measures were the proportion of correct diagnoses of neurodegenerative disease versus normal aging ("rough accuracy") and AD versus FTD ("exact accuracy"). Generalized linear mixed models were fit to assess whether the use of a volumetric report was associated with higher accuracy, accounting for random effects of within-rater and within-subject variability. Post hoc within-group analysis was performed with multiple comparisons correction. Residualized volumes were tested for an association with the diagnosis using ANOVA.

RESULTS: There was no statistically significant effect of the report on overall correct diagnoses. The proportion of "exact" correct diagnoses was higher with the report versus without the report for AD (0.52 versus 0.38) and FTD (0.49 versus 0.32) and lower for cognitively healthy (0.75 versus 0.89). The proportion of "rough" correct diagnoses of neurodegenerative disease was higher with the report than without the report within the AD group (0.59 versus 0.41), and it was similar within the FTD group (0.66 versus 0.63). Post hoc within-group analysis suggested that the report increased the accuracy in AD (OR = 2.77) and decreased the accuracy in cognitively healthy (OR = 0.25). Residualized hippocampal volumes were smaller in AD (mean difference −1.8; multiple comparisons correction, −2.8 to −0.8; $P < .001$) and FTD (mean difference −1.2; multiple comparisons correction, −2.2 to −0.1; $P = .02$) compared with cognitively healthy.

CONCLUSIONS: The availability of a brain volumetric report did not improve neuroradiologists' accuracy over visual assessment in diagnosing AD or FTD in this limited sample. Post hoc analysis suggested that the report may have biased readers incorrectly toward a diagnosis of neurodegeneration in cognitively healthy adults.

ABBREVIATIONS: AD = Alzheimer disease; CN = cognitively healthy; FTD = frontotemporal dementia

Dementia is an important and growing health care problem. In Alzheimer disease (AD) and other neurodegenerative disorders, protein misfolding leads to synaptic dysfunction and eventual neuronal loss, reflected as brain atrophy. Although

atrophy manifests late in the course of disease,¹ structural MRI is one of the most widely used diagnostic tools in the work-up of dementia.²

Brain volume changes occur in aging individuals who are cognitively healthy (CN),³ including those with subjective memory symptoms, though to a lesser degree than those with a neurodegenerative process.^{4,5} Distinguishing brain volume loss due to typical aging from a neurodegenerative process is difficult, especially early in the disease when novel treatments are thought to be potentially effective.⁶ Neuronal cell loss in AD occurs initially in the medial temporal lobes, with posterior temporal, and parietal atrophy occurring later. In contrast, the frontal and anterior temporal lobes are preferentially affected in frontotemporal

Received March 2, 2024; accepted after revision June 15.

From the Departments of Radiology (J.T., M.F.L., S.D., J.M.H., V.A.P., W.C., D.Z., A.A.T.) and Neurology (P.P., B.M.B.), University of Colorado School of Medicine, Aurora, Colorado; Center for Innovative Design and Analysis (J.P.), University of Colorado School of Public Health, Aurora, Colorado; and icometrix (B.S.), Leuven, Belgium.

Please address correspondence to Jody Tanabe, MD, 12400 E. 17th Ave L954, University of Colorado Anschutz Medical Campus, Department of Radiology, Aurora, CO 80045; e-mail: jody.tanabe@cuanschutz.edu; @jodytanabe; @AsheshThaker; @CUNeurorad

<http://dx.doi.org/10.3174/ajnr.A8406>

SUMMARY

PREVIOUS LITERATURE: Automated volumetric software analysis programs have been used in dementia research for decades. These studies have consistently demonstrated regional brain and hippocampal volume loss in AD and other neurodegenerative disorders compared with controls. However, few studies have investigated the effect of the quantitative report on neuroradiologists' ability to diagnose neurodegenerative disease compared with controls and distinguish AD from frontotemporal dementia. Furthermore, many previous studies have been conducted in research settings rather than a general clinical setting.

KEY FINDINGS: Compared with visual assessment, the availability of an automated volumetric report did not improve the overall accuracy of 4 neuroradiologists in correctly diagnosing patients as having AD or frontotemporal dementia or being cognitively intact. Post hoc analysis suggested that the report increased the accuracy in patients with AD but decreased the accuracy in the cognitively intact group.

KNOWLEDGE ADVANCEMENT: Clinically approved automated quantitative volumetric software programs for neurodegeneration may improve our understanding of diseases, but at present, the utility is questionable for dementia diagnosis in a clinical setting.

dementia (FTD), though there is heterogeneity in patterns and progression.⁷ Such atrophy patterns may suggest a diagnosis of AD versus FTD, but qualitative assessment of volume loss has been shown to be unreliable, with agreement scores reported in the 35%–70% range.⁸ Furthermore, because age-related volume loss and the effects of age on the rate and location of atrophy across diagnoses vary widely,^{9,10} normative reference databases are needed to compare volumes for a given person.

Automated quantitative volumetric analysis software programs have the potential to increase precision, decrease subjectivity, and provide a normative reference.¹¹ Volumetric differences among AD, healthy aging, and other dementias are well-established in research settings using these programs.^{12–14} However, the diagnostic utility of quantitative volumetric software in a general clinical setting is uncertain,¹⁵ particularly when encountering the common MRI brain referral indication of “memory impairment.” This study sought to determine if, compared with visual assessment alone, the availability of a clinically approved quantitative volumetric report would improve the neuroradiologist's ability to correctly diagnose AD, FTD, or CN patients drawn from a large health care system.

MATERIALS AND METHODS

The study was approved by the Colorado Multiple Institutional Review Board following expedited review.

Subjects

Subjects were identified by first screening the UHealth electronic health record between 2016 and 2021, followed by chart review. Multiple ICD-10 codes (<https://www.cdc.gov/nchs/icd/icd-10-cm/index.html>) and Boolean logic (Epic TriNetX tool; <https://www.umassmed.edu/research-informatics/resources/self-service-tools/trinetx>) were used to screen for 3 groups: AD (G30), FTD (G31.0), and other specified cognitive deficits (R41.84). Exclusion filters were neoplasm (D49.6, C71, C79.31), cerebral infarction (I63), and intracranial injury (S06). Current Procedural Terminology codes were then filtered for brain MRI within 1 year before or after an established clinical diagnosis. This process yielded 626 patients, of whom 207 were removed due to an inadequate T1-weighted anatomic MRI (ie no 3D volume acquisition, slice

thickness >1.2 mm, slice gap, contrast, and motion). The remaining 419 patient charts were reviewed by 2 authors (M.F.L. and S.D.) to verify that the diagnosis of AD or FTD was made by a neurologist, behavioral neurologist, or neurology advanced practice provider and to exclude patients with unrelated diagnoses such as MS or epilepsy. Patients with mild cognitive impairment were also excluded as a heterogeneous population,¹⁶ confounding comparison among established AD, FTD, and CN.

Three hundred sixteen were excluded after chart review (Fig 1). Ten patients with FTD met the inclusion/exclusion criteria from the electronic health record search. This group was enriched with 7 additional eligible patients from the behavioral neurology clinic for a total of 17 patients with FTD. Because the diagnosis of FTD is difficult, 1 author who is an FTD expert (P.P.) conducted an additional review of all 17 patients deemed to have FTD to ensure an accurate diagnosis. A control cohort (CN) was selected of patients who reported subjective memory symptoms without abnormal findings on neurocognitive testing or a dementia clinical diagnosis, reflecting a common indication for brain MRI. Thirty-two CN patients were initially identified through the electronic health record search, and 24 additional confirmed CN patients from the behavioral neurology clinic were identified, for a total of 56 patients. Last, to balance the groups across age and sample size, the 35 youngest CN patients and the 41 oldest patients with AD were excluded for a final sample of 60 (22 with AD, 17 with FTD, and 21 CN) (Fig 1).

Volumetric Analysis

Volumetric T1-weighted MRI studies were exported from the PACS to icobrain dm (5.7.1) for automatic report generation. The icometrix reference database is based on 1903 subjects from public databases, 6–96 years of age, 44% male.¹¹ For each participant, we collected quantitative volumes normalized to head size on 6 segmented regions (whole brain, hippocampus, frontal, temporal, occipital, and parietal). By means of the methods of Wittens et al,¹⁷ the median regional volumes from the icometrix reference data set were subtracted from each participant's observed brain volume, and these “residualized” volumes were recorded for analysis. The median reference volumes in the icometrix data set account for age and sex.

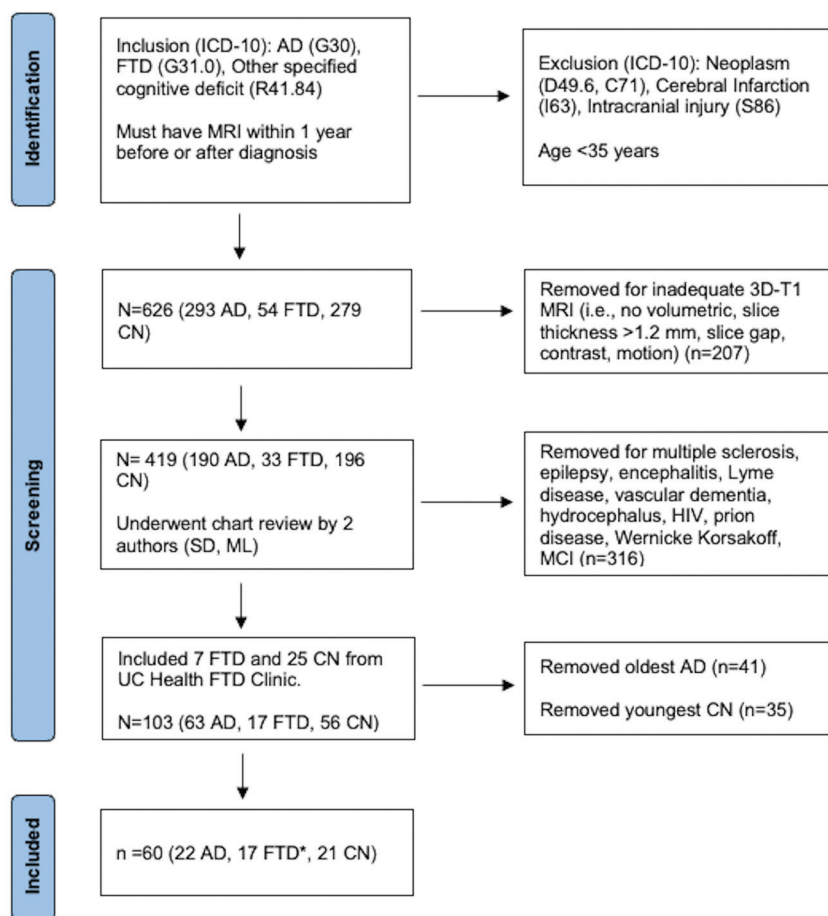


FIG 1. Patient-selection procedure. UCHHealth indicates the large health system from which the data is drawn.

Qualitative Rating

Four fellowship-trained neuroradiologists with an average of 11 years post-training experience rated T1-weighted MRI scans for each patient with and without the icometrix report. Neuroradiologists were blinded to the diagnosis but not age and sex, and none had specific expertise in dementia or dementia rating scales. Neuroradiologists first rated the patients in the study as either normal for age (ie, CN) or having a neurodegenerative disease, and this rating was defined as “rough” accuracy. If rating the study as neurodegenerative, they decided between a diagnosis of AD or FTD, and this was defined as “exact” accuracy. Studies were reviewed in 2 sessions, which were >2 weeks apart on average. At the first session, one-half of the cases were reviewed with the report and one-half without the report, and these were switched at the second session. Raters were not informed of the proportion of AD, FTD, or CN within the total sample.

Statistical Analysis

Qualitative Assessment. A generalized linear mixed model was fit to assess whether the use of the volumetric report was associated with improved diagnostic accuracy. Separate models were

fit for rough and exact accuracy. The model was fit with a binomial link to accommodate the dichotomous response (ie, correct or incorrect diagnosis), and crossed random intercepts were modeled to account for within-rater and within-patient variance. Two fixed effects were modeled: icometrix report usage (yes or no) and true patient diagnosis (AD/FTD/CN). The effect of interest is the icometrix report usage fixed effect. The 2 primary end points were exact accuracy, defined as correctly diagnosing CN, FTD, or AD, and rough accuracy, defined as correctly diagnosing CN or neurodegenerative disease. Coefficient estimates and confidence intervals are the change in the odds of correct diagnosis associated with the use of the report.

Post Hoc Within-Diagnosis Analysis.

To further understand how the report affected the radiologists’ accuracy, we assessed whether accuracy was associated with the use of the report within-diagnosis—that is, within the AD group (or CN or FTD), what were the odds that the report would be associated with radiologists’ accuracy? Generalized linear models were fit as above, with the additional inclusion of an interaction term of diagnosis with icometrix usage. Because this was a post hoc analysis, *P* values were adjusted with a Bonferroni multiple testing correction for 3 (reflect-

ing the 3 diagnosis categories). Reported *P* values have been Bonferroni-adjusted and thus can be compared directly with the nominal type I error rate.

Quantitative Assessment. Residualized volumes were assessed for an association with diagnosis (AD, FTD, CN) using ANOVA. In the event of a significant association, pair-wise group differences were assessed using the Tukey pair-wise correction for multiple comparisons.

RESULTS

Patient demographics are shown in Table 1. The FTD group was younger than those in the CN and AD groups, though older than those typical for the FTD group. Fewer patients had hypertension and diabetes in the FTD compared with the AD or CN group.

Qualitative Ratings

Figure 2 shows the proportion of correct exact and rough diagnoses with and without the report. The proportion of correct exact diagnoses was higher with the report compared with without the report for AD (0.52 versus 0.38) and FTD (0.49 versus 0.32). In contrast, the proportion of correct exact diagnoses was lower

with the report compared with without the report for CN (0.75 versus 0.89). The proportion of correct rough diagnoses was higher with the report compared with without the report for AD (0.59 versus 0.41) and similar for FTD (0.66 versus 0.63). When we modeled all groups together, there was no statistically significant effect of the report on exact (OR, 1.37, $P = .17$) or rough (OR, 1.13, $P = .55$) diagnoses.

Post Hoc Within-Group Analysis

Figure 3 shows that the report effect on correct exact and rough diagnoses was significantly worse in the CN group (exact OR, 0.26, $P = .02$; rough OR, 0.25, $P = .02$). The report effect on correct rough accuracy was significantly better in the AD group (OR, 2.77; $P = .02$)—that is, the proportion of subjects in the AD group receiving a diagnosis of AD or FTD was higher with the report.

Quantitative Volumetrics

There was a significant group difference in residualized hippocampal volume ($F = 9.91$, $P < .001$). Compared with CN individuals, residualized hippocampal volumes were lower in AD (mean difference, -1.8 ; 95% CI, -2.8 to -0.8 ; $P < .001$) and FTD (mean difference, -1.2 ; 95% CI, -2.2 to -0.1 ; $P < .02$). There was a trend toward a group difference in residualized whole-brain volume ($F = 3.10$, $P = .053$). Compared with CN individuals, whole brain volume was lower in AD but was not statistically significant (mean difference, -73.4 ; 95% CI, -150.2 – 3.4 ; $P = .06$). There were no other significant pair-wise group differences in

residualized volumes (Table 2 and Fig 4) and likewise no other significant group associations via the F test.

DISCUSSION

The availability of a quantitative report of normative brain volumes did not significantly change expert neuroradiologists' accuracy in diagnosing patients with AD, FTD, or CN, compared with qualitative visual assessment, though there was some variability based on the patient group. Specifically, post hoc analysis showed that the report increased the proportion of correct diagnoses in patient groups but decreased the proportion of correct CN diagnoses, suggesting that a quantitative report may have biased radiologists toward a disease diagnosis. Compared with CN individuals, patients with AD and FTD had lower hippocampal volumes, consistent with the literature and known biologic underpinnings.

Assessing global and regional brain volume is essential to radiologists' search patterns and warrants specific comment when screening for neurodegenerative diseases. However, visual evaluation is subjective, inconsistent, and has limited predictive ability.¹⁸ Visual rating scales can improve consistency,^{10,19–22} though the type of scale can affect agreement scores. For example, a 5-point scale had 37% agreement compared with a 2-point scale, which had 70% agreement for medial temporal lobe atrophy.⁸ Rating scales have not been widely adopted in clinical practice.

Automated volumetric software is increasingly implemented in clinical practice as more programs receive FDA approval. The sensitivity of automated programs is comparable with visual assessment²³ and manual tracings¹⁴ in distinguishing those with AD from controls. Volumetric changes spread over a 3D structure are difficult to visually detect even at moderate levels (10%–15%), highlighting the potential value of automated volumetrics.²⁴ Automated programs have consistently detected regional volume

Table 1: Demographic and group characteristics^a

	AD (n = 22)	FTD (n = 17)	CN (n = 21)
Age (yr) ^b	71.4 (4.6) [61, 78]	68.2 (10.0) [46, 82]	75.0 (5.1) [68, 85]
Male sex (%)	6 (27%)	9 (53%)	13 (62%)
Hypertension ^b	13 (59%)	2 (12%)	10 (48%)
Diabetes	5 (23%)	1 (6%)	5 (24%)
MOCA ^b	15.2 (7.4)	18.4 (6.6)	25.8 (2.8)

Note:—MOCA indicates Montreal Cognitive Assessment.

^a Data are mean (SD) [Min, Max].

^b $P < .01$.

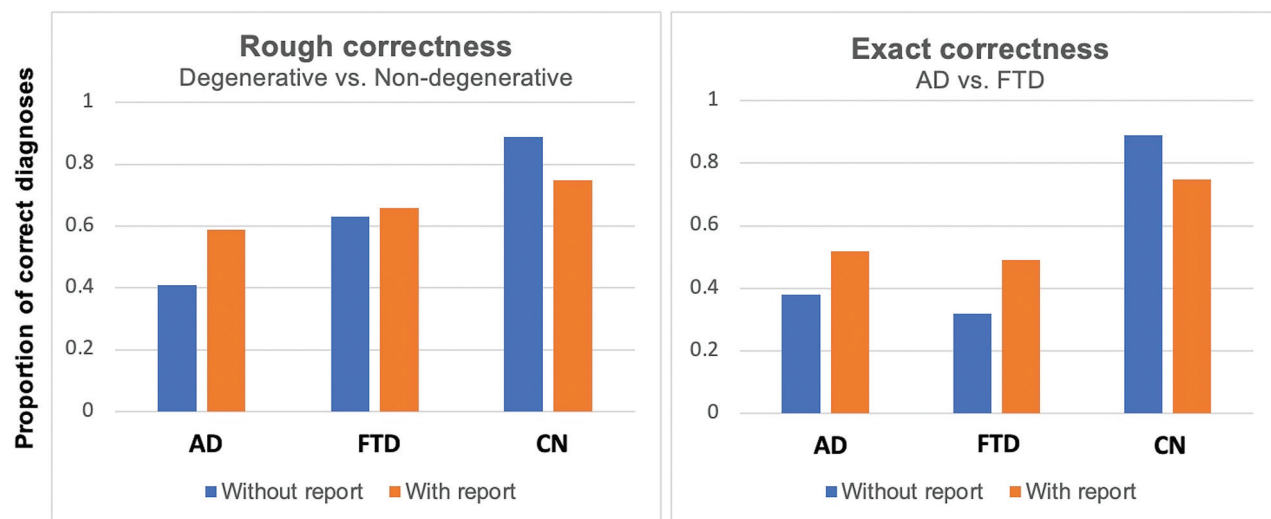


FIG 2. Proportion of neuroradiologists' exact and rough correct diagnoses with and without a quantitative report.

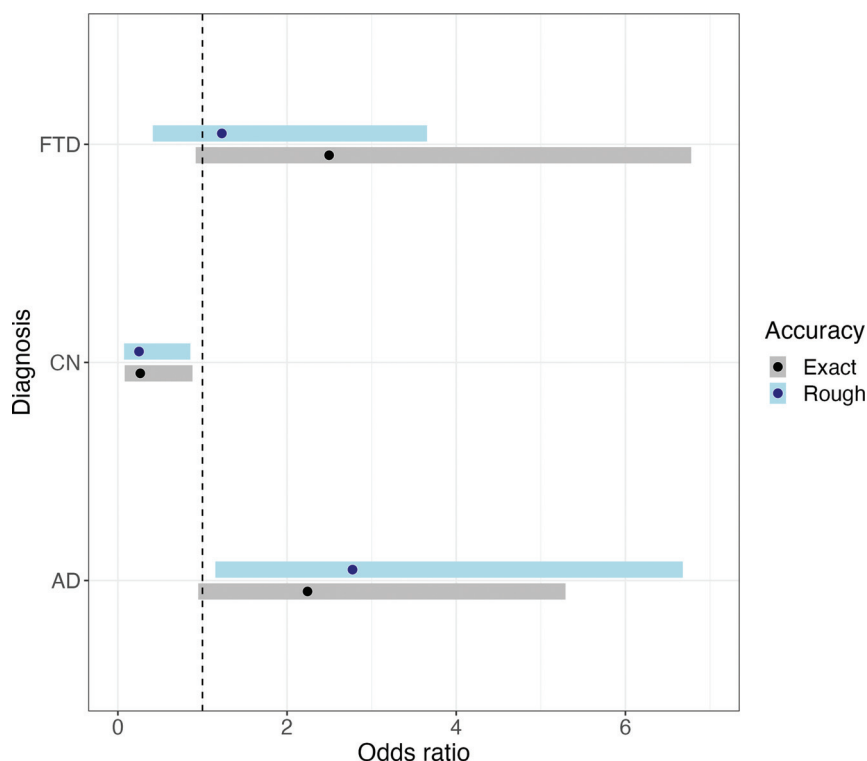


FIG 3. Plot of coefficient estimates (ORs) and 95% CIs for exact and rough correct diagnoses within group in post hoc analysis. ORs >1 indicate that icometrix improves diagnostic accuracy.

Table 2: Mean group differences, 95% CIs, and P values for residualized volumes

	Difference (AD-CN)	Lower 95% CN	Upper 95% CN	P Value
AD vs CN				
Whole brain	-73.4	-150.2	3.4	.06
Hippocampus	-1.8	-2.8	-0.8	<.001
Frontal	-1.8	-20.6	17.0	NS
Parietal	-6.1	-18.5	6.3	NS
Temporal	-10.5	-23.5	2.5	NS
Occipital	-2.6	-8.1	2.8	NS
FTD vs CN				
Whole brain	-66.4	-148.5	15.8	NS
Hippocampus	-1.2	-2.2	-0.1	0.02
Frontal	-9.5	-29.6	10.6	NS
Parietal	-0.4	-13.6	12.9	NS
Temporal	-6.5	-20.4	7.4	NS
Occipital	0.0	-5.9	5.8	NS
FTD vs AD				
Whole brain	7.0	-74.3	88.3	NS
Hippocampus	0.6	-0.4	1.6	NS
Frontal	-7.7	-27.6	12.2	NS
Parietal	5.7	-7.4	18.8	NS
Temporal	4.0	-9.7	17.8	NS
Occipital	2.6	-3.2	8.3	NS

Note:—NS indicates not significant.

differences between AD and normal aging in research settings.^{13,14,25,26} Fewer studies have focused on the effect of an automated report on diagnosis in general clinical settings.

Our finding that the report was associated with a higher proportion of correct AD diagnoses without a change in overall accuracy is partially consistent with findings in previous studies.

The study most like ours found that a quantitative report increased sensitivity, but not specificity or accuracy, for diagnosing AD versus controls. Similar to our findings, there was no significant difference in sensitivity, specificity, or accuracy in diagnosing FTD versus controls.¹⁵ Another study reported modest increases in the proportion of correct diagnoses with a report versus without one (73.5% versus 77.4% and 77.4% versus 81.1%).²⁷ Using voxel-wise color-coded maps relative to normative data,²⁷⁻²⁹ 1 study found that 1 of 2 radiologists increased diagnostic accuracy for neurodegenerative disease versus controls, and both radiologists increased their accuracy for diagnosing AD, FTD, posterior cortical atrophy, and semantic dementia.¹² Chagué et al²⁸ used artificial intelligence-generated weight maps to help readers distinguish among 4 groups: early-onset AD, late-onset AD, FTD, and patients with depression. The results were mixed in that the maps improved the radiologist's ability to distinguish those with early-onset AD

from patients with depression but did not improve the accuracy for other pair-wise group comparisons.

Rater experience might affect whether reports influence diagnostic accuracy. In a UK study, among consultants, registrars, and nonclinicians, paradoxically, only the consultant's accuracy increased with the report.¹⁵ Wibawa et al²⁹ tested the effect of volumetric reports on interrater agreement in general radiologists, neuroradiologists, and psychiatrists. Across specialists, the interrater agreement improved significantly for assessing the frontal and temporal lobes but not for the parietal lobe and hippocampal atrophy. Within a specialty, the report did not improve interrater agreement but did improve agreement overall, suggesting that reports increase overall consistency. Hedderich et al²⁷ found that the report improved classification by radiologists but not neurologists specializing in dementia. One explanation for improvement with report availability in radiologists compared with less-experienced readers is that experienced radiologists are biased against calling atrophy because they see a large range in volumes in daily practice. Although we did not test the effect of experience, evidence suggests that experience has no consistent effect on the ability for a report to alter diagnostic accuracy.

The odds of a correct diagnosis were worse with the report in the CN group; however, the highest proportion of correct diagnoses (85%) was in the CN group without the report. It is likely that raters had a high pretest probability of choosing CN because they were not informed of the proportion of each group. Low normative volumes would be expected to "nudge" radiologists to diagnose disease, resulting in a higher proportion of correct AD/FTD

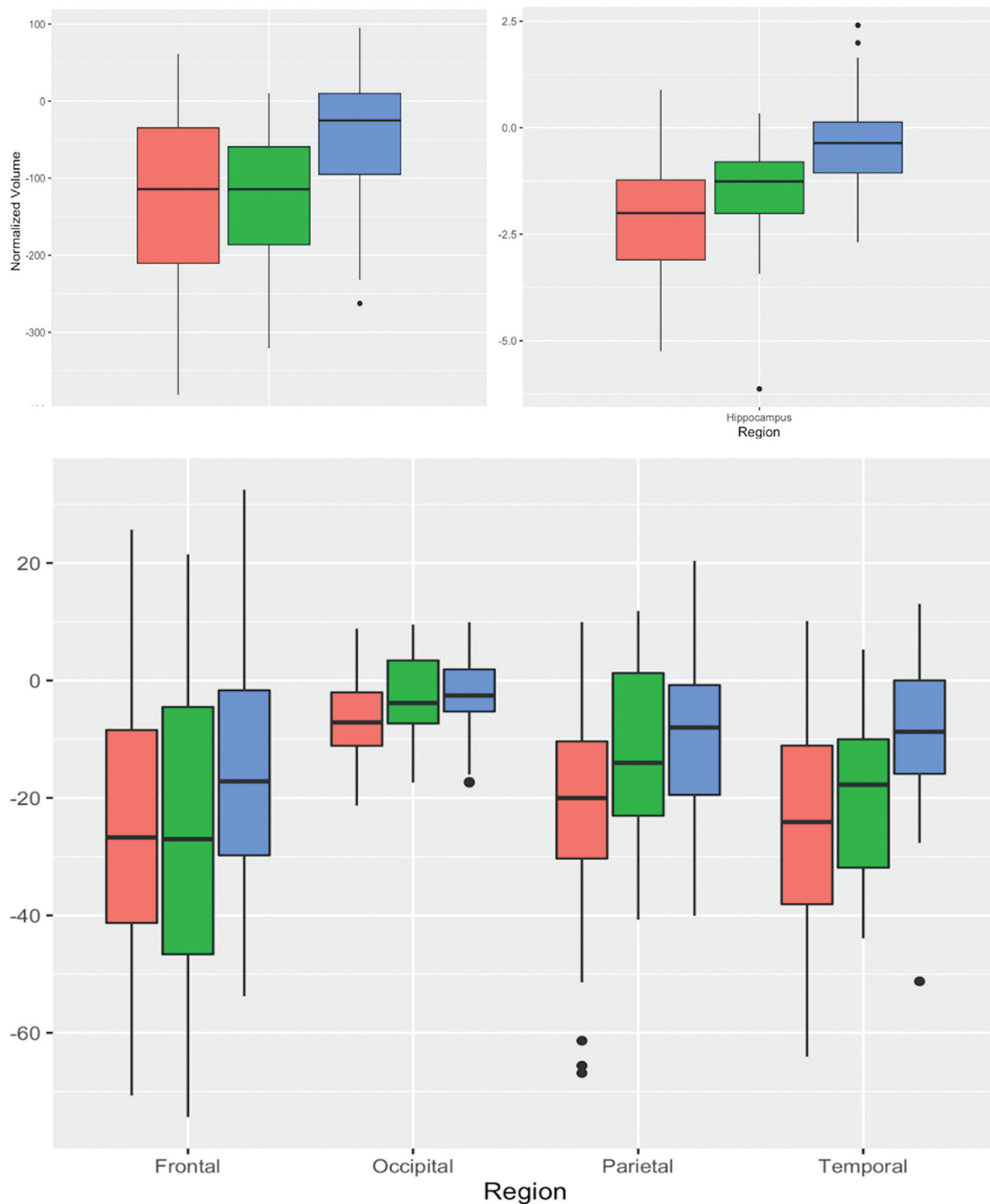


FIG 4. Boxplots of residualized volumes (blue = CN, green = FTD, red = AD). Compared with CN, both AD and FTD have significantly smaller residualized hippocampal volumes. A score of zero means there is no difference between patients' volumes and the reference data set.

diagnoses and a lower proportion of correct CN diagnoses. We did not have CSF or PET pathologic markers of Alzheimer disease pathology in any group; thus, it is possible that the CN group may have had underlying pathology and associated atrophy that had not yet clinically manifested.

We were surprised to find no difference in residualized frontal lobe volumes in those with FTD compared with CN individuals or those with AD. One reason may be that the mean age of our FTD sample was 68 years, noting that the mean age of onset of behavioral-variant FTD was reported at 58 years.³⁰ A simple lobar

imaging biomarker for FTD is probably unrealistic because FTD is exceedingly heterogeneous.^{26,31,32} Bruun et al³¹ investigated 3 indices of asymmetry and found that an anterior-posterior index differentiated behavioral-variant FTD from other dementias, while a left-right index best distinguished behavioral-variant FTD from primary-progressive aphasia. In a large sample of 1213 patients, modest sensitivities of 59% to 82% underscore the challenge of finding an imaging biomarker for FTD. Even within a subtype of FTD, there may be different atrophy patterns.^{32,33}

Limitations

The study sample size of 60 was small due to careful chart screening, but the size was in keeping with that in similar studies.^{23,25,34} Among 600 patients screened, one-third were excluded due to an inadequate MRI, which could be a source of bias. Excluding patients with mild cognitive impairment and other neurodegenerative diseases does not mirror a typical clinical setting; however, we sought to minimize the possibility of a misdiagnosis. While some selection bias is inherent to a retrospective cross-sectional design, our study was successful in identifying 3 distinct groups with clinical diagnoses with baseline demographic characteristics suitable for comparison. Residualized volumes in our study were lower than those in the reference for all groups, suggesting potential systematic bias. However, the effect would be mitigated because the analysis was between groups. Because this was a cross-sectional study, the risk for future decline in CN individuals is possible, but this risk is mitigated by patients with mild cognitive impairment being excluded and MRI being acquired within 1 year of diagnosis. Diabetes and hypertension have been associated with brain volume change and cognitive impairment;³⁵ however, the frequency of diabetes and hypertension was similar in the AD and CN groups.

Another limitation is that the diagnosis of AD and FTD was purely clinical without confirmatory biomarkers of amyloid or τ pathology.³⁶ Structural MRI captures only neurodegeneration, which, as a single biomarker in the AT (N) biomarker classification, includes both AD and non-AD pathologic change. Thus, radiologists' ability to correctly classify AD compared with CN may be confounded by individuals in our study with non-AD pathologic change. As access to amyloid and τ biomarkers improves, further research may clarify the role of structural MRI in supporting dementia diagnoses. At present, the utility of quantitative structural MRI tools is questionable for a dementia diagnosis in a clinical setting.

CONCLUSIONS

The availability of a quantitative report of brain volumes did not change the proportion of correct diagnoses among a small sample of patients with AD, FTD, and CN drawn from a single health care system. Post hoc analyses suggest that the report may lower the threshold of experienced neuroradiologists suggesting a diagnosis of neurodegenerative disease. Possible sources of bias include pretest probability, rater experience, and the accuracy of the reference database.

Disclosure forms provided by the authors are available with the full text and PDF of this article at www.ajnr.org.

REFERENCES

1. Jack CR Jr, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* 2010;9:119–28 [CrossRef Medline](#)
2. Chouliaras L, O'Brien JT. The use of neuroimaging techniques in the early and differential diagnosis of dementia. *Mol Psychiatry* 2023;28:4084–97 [CrossRef Medline](#)
3. Fujita S, Mori S, Onda K, et al. Characterization of brain volume changes in aging individuals with normal cognition using serial magnetic resonance imaging. *JAMA Netw Open* 2023;6:e2318153 [CrossRef Medline](#)
4. Fyock CA, Hampstead BM. Comparing the relationship between subjective memory complaints, objective memory performance, and medial temporal lobe volumes in patients with mild cognitive impairment. *Alzheimers Dement (Amst)* 2015;1:242–48 [CrossRef Medline](#)
5. Kawagoe T, Onoda K, Yamaguchi S. Subjective memory complaints are associated with altered resting-state functional connectivity but not structural atrophy. *Neuroimage Clin* 2019;21:101675 [CrossRef Medline](#)
6. Passeri E, Elkhoury K, Morsink M, et al. Alzheimer's disease: treatment strategies and their limitations. *Int J Mol Sci* 2022;23:13954 [CrossRef Medline](#)
7. Boeve BF, Boxer AL, Kumfor F, et al. Advances and controversies in frontotemporal dementia: diagnosis, biomarkers, and therapeutic considerations. *Lancet Neurol* 2022;21:258–72 [CrossRef Medline](#)
8. Scheltens P, Launer LJ, Barkhof F, et al. Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. *J Neurol* 1995;242:557–60 [CrossRef Medline](#)
9. Rhodius-Meester HF, Benedictus MR, Wattjes MP, et al. MRI visual ratings of brain atrophy and white matter hyperintensities across the spectrum of cognitive decline are differently affected by age and diagnosis. *Front Aging Neurosci* 2017;9:117 [CrossRef Medline](#)
10. Koedam EL, Lehmann M, van der Flier WM, et al. Visual assessment of posterior atrophy development of a MRI rating scale. *Eur Radiol* 2011;21:2618–25 [CrossRef Medline](#)
11. Pemberton HG, Zaki LA, Goodkin O, et al. Technical and clinical validation of commercial automated volumetric MRI tools for dementia diagnosis—a systematic review. *Neuroradiology* 2021;63:1773–89 [CrossRef Medline](#)
12. Hedderich DM, Dieckmeyer M, Andrisan T, et al. Normative brain volume reports may improve differential diagnosis of dementing neurodegenerative diseases in clinical practice. *Eur Radiology* 2020;30:2821–29 [CrossRef Medline](#)
13. Niemantsverdriet E, Ribbens A, Bastin C, et al. A retrospective Belgian multi-center MRI biomarker study in Alzheimer's Disease (REMEMBER). *J Alzheimers Dis* 2018;63:1509–22 [CrossRef Medline](#)
14. Brewer JB. Fully-automated volumetric MRI with normative ranges: translation to clinical practice. *Behav Neurol* 2009;21:21–28 [CrossRef Medline](#)
15. Pemberton HG, Goodkin O, Prados F, et al; Alzheimer's Disease Neuroimaging Initiative. Automated quantitative MRI volumetry reports support diagnostic interpretation in dementia: a multi-rater, clinical accuracy study. *Eur Radiol* 2021;31:5312–23 [CrossRef Medline](#)
16. Dunne RA, Aarsland D, O'Brien JT, et al. Mild cognitive impairment: the Manchester consensus. *Age Ageing* 2021;50:72–80 [CrossRef Medline](#)
17. Wittens MM, Sima DM, Houbrechts R, et al. Diagnostic performance of automated MRI volumetry by icobrain dm for Alzheimer's disease in a clinical setting: a REMEMBER study. *J Alzheimers Dis* 2021;83:623–39 [CrossRef Medline](#)
18. Ten Kate M, Barkhof F, Boccardi M, et al; Geneva Task Force for the Roadmap of Alzheimer's Biomarkers. Clinical validity of medial temporal atrophy as a biomarker for Alzheimer's disease in the context of a structured 5-phase development framework. *Neurobiol Aging* 2017;52:167–82.e1 [CrossRef Medline](#)

19. Harper L, Barkhof F, Fox NC, et al. **Using visual rating to diagnose dementia: a critical evaluation of MRI atrophy scales.** *J Neurol Neurosurg Psychiatry* 2015;86:1225–33 [CrossRef Medline](#)
20. Harper L, Fumagalli GG, Barkhof F, et al. **MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases.** *Brain* 2016;139:1211–25 [CrossRef Medline](#)
21. Scheltens P, Leys D, Barkhof F, et al. **Atrophy of medial temporal lobes on MRI in “probable” Alzheimer’s disease and normal ageing: diagnostic value and neuropsychological correlates.** *J Neurol Neurosurg Psychiatry* 1992;55:967–72 [CrossRef Medline](#)
22. Sarria-Estrada S, Acevedo C, Mitjana R, et al. **Reproducibility of qualitative assessments of temporal lobe atrophy in MRI studies [Article in English and Spanish].** *Radiologia* 2015;57:225–28 [CrossRef Medline](#)
23. Min J, Moon WJ, Jeon JY, et al. **Diagnostic efficacy of structural MRI in patients with mild-to-moderate Alzheimer disease: automated volumetric assessment versus visual assessment.** *AJR Am J Roentgenol* 2017;208:617–23 [CrossRef Medline](#)
24. Ross DE, Ochs AL, DeSmit ME, et al; Alzheimer’s Disease Neuroimaging Initiative. **Man versus machine, Part 2: comparison of radiologists’ interpretations and Neuroquant measures of brain asymmetry and progressive atrophy in patients with traumatic brain injury.** *J Neuropsychiatry Clin Neurosci* 2015;27:147–52 [CrossRef Medline](#)
25. Bottino CM, Castro CC, Gomes RL, et al. **Volumetric MRI measurements can differentiate Alzheimer’s disease, mild cognitive impairment, and normal aging.** *Int Psychogeriatr* 2002;14:59–72 [CrossRef Medline](#)
26. Yu Q, Mai Y, Ruan Y, et al; Frontotemporal Lobar Degeneration Neuroimaging Initiative. **An MRI-based strategy for differentiation of frontotemporal dementia and Alzheimer’s disease.** *Alzheimers Res Ther* 2021;13:23 [CrossRef Medline](#)
27. Hedderich DM, Spiro JE, Goldhardt O, et al. **Increasing diagnostic accuracy of mild cognitive impairment due to Alzheimer’s disease by user-independent, Web-based whole-brain volumetry.** *J Alzheimers Dis* 2018;65:1459–67 [CrossRef Medline](#)
28. Chagué P, Marro B, Fadili S, et al. **Radiological classification of dementia from anatomical MRI assisted by machine learning-derived maps.** *J Neuroradiol* 2021;48:412–18 [CrossRef Medline](#)
29. Wibawa P, Matta G, Das S, et al. **Bringing psychiatrists into the picture: Automated measurement of regional MRI brain volume in patients with suspected dementia.** *Aust N Z J Psychiatry* 2021;55:799–808 [CrossRef Medline](#)
30. Pressman PS, Miller BL. **Diagnosis and management of behavioral variant frontotemporal dementia.** *Biol Psychiatry* 2014;75:574–81 [CrossRef Medline](#)
31. Bruun M, Koikkalainen J, Rhodius-Meester HF, et al. **Detecting frontotemporal dementia syndromes using MRI biomarkers.** *Neuroimage Clin* 2019;22:101711 [CrossRef Medline](#)
32. Perry RJ, Graham A, Williams G, et al. **Patterns of frontal lobe atrophy in frontotemporal dementia: a volumetric MRI study.** *Dement Geriatr Cogn Disord* 2006;22:278–87 [CrossRef Medline](#)
33. Bocchetta M, Iglesias JE, Scelsi MA, et al. **Hippocampal subfield volumetry: differential pattern of atrophy in different forms of genetic frontotemporal dementia.** *J Alzheimers Dis* 2018;64:497–504 [CrossRef Medline](#)
34. Barber R, Ballard C, McKeith IG, et al. **MRI volumetric study of dementia with Lewy bodies: a comparison with AD and vascular dementia.** *Neurology* 2000;54:1304–09 [CrossRef Medline](#)
35. Roberts RO, Knopman DS, Przybelski SA, et al. **Association of type 2 diabetes with brain atrophy and cognitive impairment.** *Neurology* 2014;82:1132–41 [CrossRef Medline](#)
36. Jack CR Jr, Bennett DA, Blennow K, et al. Contributors. **NIA-AA Research Framework: toward a biological definition of Alzheimer’s disease.** *Alzheimers Dement* 2018;14:535–62 [CrossRef Medline](#)