pISSN 2384-1095
eISSN 2384-1109

iMRI 2015;19:88-98

http://dx.doi.org/10.13104/imri.2015.19.2.88

iMRI
Investigative Magnetic Resonance Imaging

Measurement of Apparent Diffusion Coefficient Values from Diffusion-Weighted MRI: A Comparison of Manual and Semiautomatic Segmentation Methods

Seong Ho Kim[1], Seung Hong Choi[1,2], Tae Jin Yoon[1], Tae Min Kim[3], Se-Hoon Lee[3], Chul-Kee Park[4], Ji-Hoon Kim[1], Chul-Ho Sohn[1], Sung-Hye Park[5], Il Han Kim[6]

[1]Department of Radiology, Seoul National University College of Medicine, Seoul, Korea
[2]Center for Nanoparticle Research, Institute for Basic Science, and School of Chemical and Biological Engineering, Seoul National University, Seoul, Korea
[3]Department of Internal Medicine, Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea
[4]Department of Neurosurgery, Seoul National University College of Medicine, Seoul, Korea
[5]Department of Pathology, Seoul National University College of Medicine, Seoul, Korea
[6]Department of Radiation Oncology, Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea

Original Article

Received: April 6, 2015
Revised: May 4, 2015
Accepted: May 7, 2015

Correspondence to:
Seung Hong Choi, M.D., Ph.D.
Department of Radiology, Seoul National University College of Medicine, 28, Yongon-dong, Chongno-gu, Seoul 110-744, Korea.
Tel. +82-2-2072-2861
Fax. +82-2-743-6385
Email: verocay@snuh.org

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2015 Korean Society of Magnetic Resonance in Medicine (KSMRM)

Purpose: To compare the interobserver and intraobserver reliability of mean apparent diffusion coefficient (ADC) values using contrast-enhanced (CE) T1 weighted image (WI) and T2WI as structural images between manual and semiautomatic segmentation methods.

Materials and Methods: Between January 2011 and May 2013, 28 patients who underwent brain MR with diffusion weighted image (DWI) and were pathologically confirmed as having glioblastoma participated in our study. The ADC values were measured twice in manual and semiautomatic segmentation methods using CE-T1WI and T2WI as structural images to obtain interobserver and intraobserver reliability. Moreover, intraobserver reliabilities of the different segmentation methods were assessed after subgrouping of the patients based on the MR findings.

Results: Interobserver and intraobserver reliabilities were high in both manual and semiautomatic segmentation methods on CE-T1WI-based evaluation, while interobserver reliability on T2WI-based evaluation was not high enough to be used in a clinical context. The intraobserver reliability was particularly lower with the T2WI-based semiautomatic segmentation method in the subgroups with involved lobes ≤ 2, with partially demarcated tumor borders, poorly demarcated inner margins of the necrotic portion, and with perilesional edema.

Conclusion: Both the manual and semiautomatic segmentation methods on CE-T1WI-based evaluation were clinically acceptable in the measurement of mean ADC values with high interobserver and intraobserver reliabilities.

Keywords: Glioblastoma; ADC; Segmentation; Manual; Semiautomatic

INTRODUCTION

Glioblastoma multiforme (GBM) is one of the most lethal gliomas in adults, and it results in more years of life lost than other brain tumors (1). Because the therapeutic agent can differ according to the tumor grade, it is critical to maintain the reliability and reproducibility of tumor grading to provide effective therapeutic agents for patients with gliomas (2, 3). Various assessments of the differentiation of tumor grade have been proposed for use in planning treatment methods (4). It remains doubtful whether tumor grade can be evaluated with sufficient specificity on diffusion-weighted (DW) MRI with quantitative apparent diffusion coefficients (ADCs) (5-7). However, DWI with ADC is reportedly useful in predicting the response to treatment with bevacizumab for newly diagnosed or recurrent GBM (4, 8, 9).

Several studies have addressed the segmentation of GBM with DWI and ADC maps that facilitate the prediction of tumor response after treatment. Fuzzy clustering and knowledge-based analysis were proposed in several early studies (10, 11). More recently, studies have explored voxel-based classification methods (12-14). Most of the aforementioned studies performed segmentation methods using multiple sequences, such as T1 weighted image (WI) and T2WI. In recent years, the semiautomatic method of drawing the region of interest (ROI) on contrast-enhanced T1WI (CE-T1WI) and thereafter mapping the tumor contour on the ADC map was proposed for the segmentation of GBM (8). More methods of computer-aided diagnosis in GBM have been proposed and are currently used in research (12, 13). However, it is still unclear which method is more reproducible and more useful in a clinical context in the characterization of GBM between manual and semiautomatic segmentation methods using DWI and ADC maps.

Hence, the purpose of our study was to compare interobserver and intraobserver reproducibility between the semiautomatic segmentation method and the manual segmentation method using CE-T1WI and T2WI as structural images by calculating the mean ADC values, which were derived from DWI in GBM. Furthermore, our aim was to compare intraobserver reproducibility between 2 different segmentation methods among the subgroups, which were divided according to the MR imaging findings.

MATERIALS AND METHODS

Patients

The Institutional Review Board of Seoul National University Hospital approved this study, and informed consent was waived because our study was retrospectively designed. Between January 2011 to May 2013, 88 patients underwent brain MR with DW MR imaging in our institution, with pathologic confirmation of grade IV GBM according to the World Health Organization (WHO) criteria. Twenty-eight (18 men, mean age 54.2 years old, age range 16-82 years old; 10 women, mean age 49 years old, age range 20-72 years old) of the 88 patients were selected after performing power analysis. All 28 patients who were enrolled in our study underwent MR imaging with DWI before treatment (Fig. 1).

MR Technique

All 28 patients were scanned with the same 3 Tesla (T) MR unit (Verio, Siemens Healthcare, Erlangen, Germany) with a 32-channel head coil. MR images were obtained on the axial plane with a field of view (FOV) of 240 x 240 mm, adjusted for each patient. The baseline MR images included transverse T2WI, using a turbo spin-echo (TSE) sequence (repetition time [TR], 5160 ms; echo time [TE], 91 ms; flip angle [FA], 130°; matrix, 640 × 540; section thickness, 5 mm; and number of excitations [NEX], 3), and multi-planar reconstructed transverse, coronal T1WI, with a sagittal three-dimensional magnetization-prepared rapid acquisition gradient echo (3D-MPRAGE) sequence (TR, 1500 ms; TE, 1.9 ms; FA, 9°; matrix, 256 × 232; section thickness, 1 mm; and NEX, 1). Echo-planar DW MR imaging of b = 0 and 1000 sec/mm^2 (TR, 6000 ms; TE, 63 ms; FA, 90°; matrix, 160 × 160; section thickness, 5 mm; intersection gap, 1 mm; sections, 25; voxel size, 1.5 × 1.5 × 5.0 mm^3; and bandwidth, 1953 Hz) was also acquired from the axial images. Then, an ADC map was calculated on a voxel-by-voxel basis with the software incorporated into the MR unit. CE-T1WI was acquired after the administration of gadobutrol (Gadovist®, Gd-BT-DO3A, Bayer Schering Pharma, Berlin, Germany) at a dose of 0.1 mmol/kg and an injection rate of 2 mL/sec.

Image Processing and Analysis

Histogram of the ADC values were obtained from the semiautomatic and manual segmentation methods, using structural images of CE-T1WI and T2WI, using commercialized software (Nordic TumorEx and Nordic

ICE, NordicNeuroLab, Bergen, Norway). Histograms of the ADC values are assessed with multiple pixels, using the automatically coregistered images between the standard ADC map and the structural images of CE-T1WI or T2WI (15, 16). The coregistration between the structural images and parametric maps (color overlay) was accomplished automatically using mutual information based on an algorithm that facilitated the search for an optimal rigid transformation aligning the 2 data sets (16, 17). The ADC values were obtained in all cases independently by the 2 radiologists (S.H.K. and S.H.C. with 3 and 13 years of experience, respectively). They determined the extent of viable tumor using both the manual and semiautomatic segmentation methods on CE-T1WI- and T2WI-based evaluations, respectively. Then, all of the ADC values were obtained within the segmented tumor. The mean ADC value was calculated with electronic software (Excel, Microsoft). The measurement of ADC values was performed twice with the manual segmentation method and semiautomatic segmentation method using CE-T1WI and T2WI for structural images, respectively, over a period of 6 months. A total of 8 time measurements were performed for each case, with an interval of at least 2 weeks between the different segmentation methods and between different structural images to reduce the recall bias (e.g., twice with the manual segmentation method on CE-T1WI for structural images, twice with the manual segmentation method on T2WI for structural images, twice with the semiautomatic segmentation method on CE-T1WI for structural images, and twice with the semiautomatic segmentation method on T2WI for structural images).

After measurement of the ADC values, the characteristics of the MR findings were evaluated in all of the patients based on the consensus of the same 2 radiologists. Moreover, the same 2 radiologists consensually selected 1 of 2 ADC values for further assessment of subtracted value between 1st and 2nd mean ADC values to evaluate the intraobserver reliability among the various subgroups that were divided according to the MR imaging findings. The subtracted mean ADC value was obtained independently for each segmentation method on each structural image in every patient. We considered the calculation to be more reliable when the mean difference of the 1st and 2nd ADC value was smaller and closer to zero. The reviewed features of GBM with MR imaging were as follows: tumor size, number of involved lobes, tumor border, proportion of necrosis, inner margin of necrotic portion, and perilesional edema. To obtain the tumor size, the largest diameter was measured on the axial plane of CE-T1WI. The largest diameter of the largest tumor was measured in cases

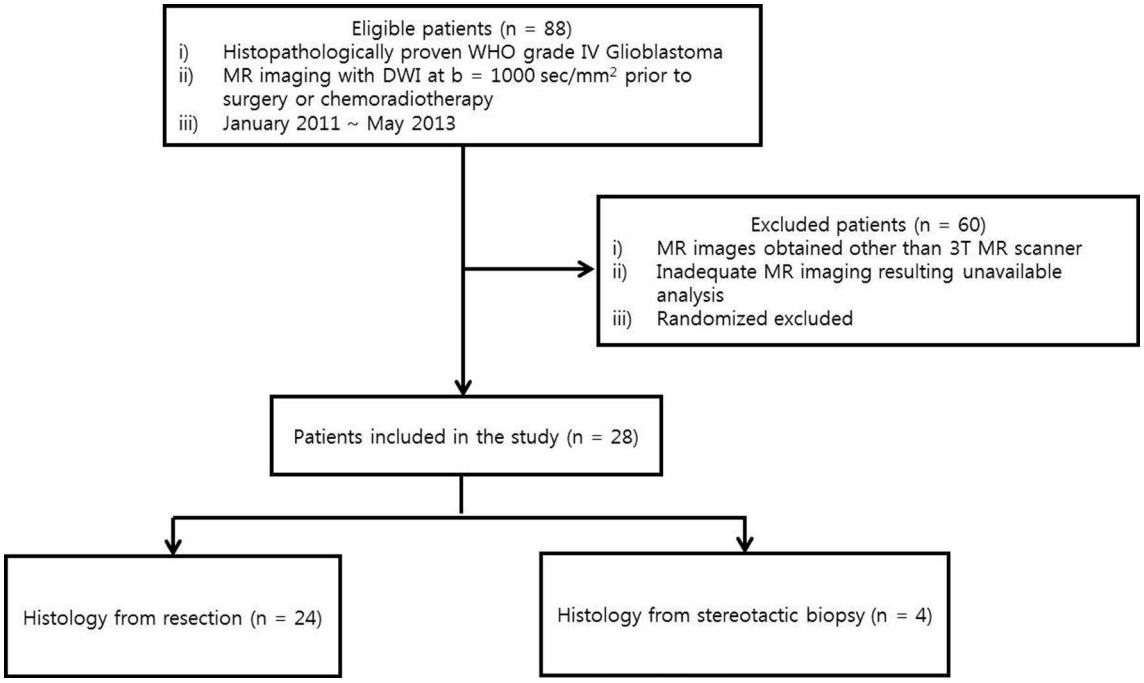

Fig. 1. Flowchart of patient selection. DWI = diffusion-weighted imaging; MR = magnetic resonance; T = tesla; WHO = World Health Organization

of multiple tumors. The number of involved lobes was classified into 2 groups i.e., group 1 with involved lobes ≤ 2; and group 2 with involved lobes > 2. The tumor border was defined as well demarcated (infiltrative border ≤ 10%), partially poorly demarcated (10% < infiltrative border < 90%), or fully poorly demarcated (infiltrative border ≥ 90%). The proportion of necrosis was divided into 3 groups i.e., predominantly solid (necrotic portion ≤ 25% of the tumor), solid and cystic (necrotic portion in the range of 25% to 75%), or predominantly cystic (necrotic portion ≥ 75% of the tumor). The inner margin of the necrotic portion was classified as well demarcated (infiltrative inner margin ≤ 25%) or poorly demarcated (infiltrative inner margin > 25%), regardless of the outer tumor margin. The outer and inner tumor margins, as well as the tumor components, were evaluated on the axial plane of CE-T1WI. Perilesional edema was considered when the peritumoral T2 signal different from the tumor increased on the axial T2WI.

A. Manual segmentation method (Fig. 2)

Coregistration between the standard ADC maps and structural images was performed for the first step using commercialized software (Nordic ICE, NordicNeuroLab, Bergen, Norway). Then, 2 observers (S.H.K and S.H.C) outlined the viable tumor on every axial plane of CE-T1WI and T2WI for the structural images to obtain the histogram of the ADC values. The histogram of the ADC values was obtained on the basis of using multiple pixels. We outlined the tumor on CE-T1WI with reference to the enhanced portion and on T2WI with reference to the signal difference, as compared with normal brain parenchyma but avoiding the cystic or necrotic portion, intralesional macrovessels, and perilesional edema. The mean ADC value was calculated manually using the obtained ADC values within the ROI on histogram. The same process was repeated to evaluate intraobserver reliability, with an interval of at least 2 weeks between measurements.

B. Semiautomatic segmentation method (Fig. 2)

The semiautomatic segmentation was performed using a different commercialized software (Nordic TumorEx, NordicNeuroLab, Bergen, Norway). The same 2 observers (S.H.K and S.H.C) manually defined the elliptical volume of interest (VOI), including the entire mass on the structural images of CE-T1WI and T2WI, the software automatically segmented the tumor components only within the defined VOI using clustering analysis. The software provides 3 to 7 clusters without overlaps, within the defined VOI. We used the mode that provided 6 clusters within the defined VOI. Two observers selected at least 1 cluster to include the entire viable portion of the mass. The software was limited in differentiating between the enhanced portion and vascular structures on CE-T1WI and automatically differentiating between perilesional edema and the tumor on T2WI, hence additional manual correction was performed to include only pure, viable tumor. Manual correction was done by drawing or erasing necessary or unnecessary portion of the viable tumor by clicking with the mouse. The major difference between semiautomatic and manual segmentation method is that basic outline of the viable tumor is determined automatically in the semiautomatic segmentation method and drawn manually using the mouse in the manual segmentation method. Thereafter, the histogram of ADC values within the VOI was obtained on the basis of multiple pixels using the Nordic ICE software. The mean ADC value was also calculated manually. The same process was independently repeated to evaluate the intraobserver reliability on CE-T1WI- and T2WI-based structural images, with an interval of at least 2 weeks between measurements.

Statistical Analysis

Statistical analysis was performed using commercially available software (MedCalc, version 12.7.1.0, MedCalc software, Mariakerke, Belgium). We assessed interobserver reliability between 2 observers and intraobserver reliability between the 1st and 2nd measurements in each segmentation method using CE-T1WI- and T2WI-based structural images, respectively. We used the intraclass correlation coefficient (ICC), coefficient of variation (CV), and Bland-Altman plot to evaluate the reproducibility. The ICC values were categorized as follows: < 0.40, poor agreement; 0.40-0.59, fair agreement; 0.60-0.74, good agreement; and ≥ 0.75, high agreement (12). The CV value was defined as 100 × the standard deviation/mean (%) (18).

The subtracted mean ADC value between the 1st and 2nd measurements in each segmentation method for the CE-T1WI- and T2WI-based structural images was used to compare the difference in intraobserver reproducibility between the segmentation methods. We considered it more reproducible when the subtracted mean of the ADC value was smaller and closer to zero. The paired t-test was used to compare the difference in intraobserver reproducibility between the segmentation methods in each subgroup, divided by the MR findings. A P-value less than 0.05 was considered as statistical significance.

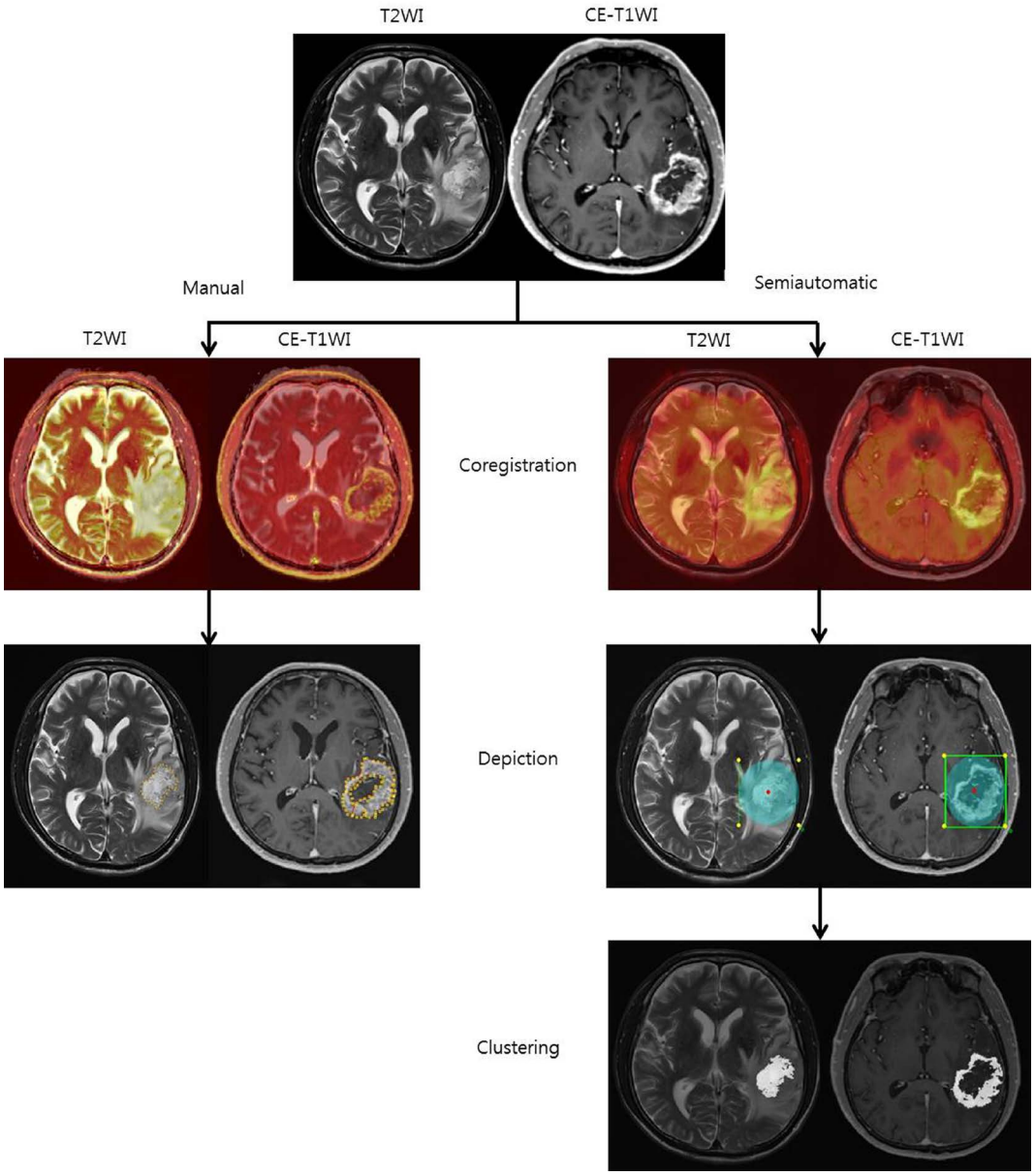

Fig. 2. Flowchart of manual and semiautomatic segmentation based on T2WI and CE-T1WI as structural images. A 70-year-old man with WHO grade IV glioblastoma confirmed via resection, underwent MR imaging with DWI before surgery or chemoradiotherapy. (Top) The axial image of T2-weighted turbo spin echo sequence demonstrated a T2 high signal intensity mass in the left temporal lobe with definite perilesional edema. The mass measured approximately 4.8 cm at its largest diameter. The reformatted axial image of CE-T1WI demonstrates a well-enhanced, solid, and cystic mass with a partially poorly demarcated outer tumor border, as well as a poorly demarcated inner margin of the necrotic portion. Intralesional macrovessels were also noted. (Left column) Coregistrations between structural images (T2WI and CE-T1WI) and the ADC map were performed using the manual segmentation method. Thereafter, the ROI was depicted manually by the reviewers on the axial planes of both T2WI and CE-T1WI. (Right column) Structural images (T2WI and CE-T1WI) and the ADC map were coregistered using the semiautomatic segmentation method. The reviewers manually defined the elliptical VOI, including the entire mass, on structural images. The software automatically segmented the tumor solely within the defined VOI using clustering analysis. Finally, the reviewers depicted appropriate combinations of clusters for tumor segmentation. ADC = apparent diffusion coefficient; CE-T1WI = contrast-enhanced T1-weighted imaging; DWI = diffusion-weighted imaging; MR = magnetic resonance; ROI = region of interest; T2WI = T2-weighted imaging; VOI = volume of interest; WHO = World Health Organization

RESULTS

Tumor Characteristics

The mean tumor size in our study was 5.1 cm. Nine cases showed mass involvement of > 2 lobes (32%), and the remaining 19 cases showed mass involvement of ≤ 2 lobes (68%). Nine cases showed well-demarcated margins (32%), 11 cases showed partially poorly demarcated margins (39%), and the remaining 8 cases showed fully poorly demarcated margins (29%). Only 4 cases showed predominantly solid features (14%); 9 cases showed solid and cystic features (32%), and 15 cases showed predominantly cystic features (54%). Moreover, 5 cases showed well-demarcated inner margins of the necrotic portion (18%), and 23 cases showed poorly demarcated inner margins of the necrotic portion (82%). Definite perilesional edema was present in 25 cases (89%).

Interobserver and Intraobserver Reliability in the Manual and Semiautomatic Segmentation Methods

Mean ADC values were listed in Table 1, which were evaluated by the 2 radiologists using manual and semiautomatic segmentation methods on CE-T1WI- and T2WI-based evaluation, respectively. The ICC and CV values of the manual and semiautomatic segmentation methods for interobserver and intraobserver reliabilities were listed in Tables 2 and 3, respectively. Interobserver reliability was not high enough to be used in a clinical context on T2WI-based evaluation, particularly with the semiautomatic segmentation method. For intraobserver reliability, the ICC

Table 1. Mean ADC Values Obtained by the 2 Radiologists Using Manual and Semiautomatic Segmentation Methods

	Radiologist 1							
	1st* Manual CE-T1WI	1st* Manual T2WI	2nd* Manual CE-T1WI	2nd* Manual T2WI	1st* Semiautomatic CE-T1WI	1st* Semiautomatic T2WI	2nd* Semiautomatic CE-T1WI	2nd* Semiautomatic T2WI
Mean	1133.74	1096.66	1132.92	1080.85	1119.80	1177.79	1116.20	1134.86
Range	786.82-1443.79	763.38-1486.38	769.19-1439.87	754.49-1457.75	744.66-1426.50	740.61-1866.52	740.62-1425.20	770.50-1709.69
	Radiologist 2							
	1st* Manual CE-T1WI	1st* Manual T2WI	2nd* Manual CE-T1WI	2nd* Manual T2WI	1st* Semiautomatic CE-T1WI	1st* Semiautomatic T2WI	2nd* Semiautomatic CE-T1WI	2nd* Semiautomatic T2WI
Mean	1081.34	1177.59	1075.87	1190.68	1102.67	1245.71	1101.21	1238.25
Range	732.97-1431.14	853.30-1489.25	732.74-1443.85	860.96-1515.33	705.65-1453.78	876.65-1582.28	738.60-1425.36	869.58-1599.60

ADC = apparent diffusion coefficient; CE-T1WI = contrast-enhanced T1-weighted imaging; T2WI = T2-weighted imaging

*1st and 2nd indicate the first and second measurement, respectively.

Table 2. Interobserver Reliability in ADC Measurement

	Manual method			
	1st* CE-T1WI	1st* T2WI	2nd* CE-T1WI	2nd* T2WI
ICC†	0.950 (0.895-0.976)	0.728 (0.496-0.864)	0.929 (0.854-0.966)	0.651 (0.377-0.821)
CV‡	4.25	9.30	5.11	10.59
	Semiautomatic segmentation method			
	1st* CE-T1WI	1st* T2WI	2nd* CE-T1WI	2nd* T2WI
ICC†	0.988 (0.975-0.995)	0.572 (0.264-0.775)	0.994 (0.987-0.997)	0.624 (0.338-0.806)
CV‡	2.05	13.52	1.44	11.84

All of the numbers in brackets indicate the 95% confidence interval.

ADC = apparent diffusion coefficient; CE-T1WI = contrast-enhanced T1-weighted imaging; CV = coefficient of variance; ICC = intraclass correlation coefficient; T2WI = T2-weighted imaging

*1st and 2nd indicate the first and second measurement, respectively.

†ICC values were categorized as follows: < 0.40, poor agreement; 0.40–0.59, fair agreement; 0.60–0.74, good agreement; and ≥ 0.75, high agreement.

‡Numbers are expressed as percentages.

value was the lowest in the semiautomatic segmentation method on T2WI-based evaluation, and the CV value was the highest among the measurements in both observers. Nevertheless, both the manual and semiautomatic segmentation methods showed high reproducibility in the intraobserver measurements using both CE-T1WI and T2WI for structural images. The Bland-Altman plot analysis also supported the aforementioned findings of high intraobserver reliability in both the manual and semiautomatic segmentation methods (Fig. 3).

Intraobserver Reliability among the Different Subgroups

We subgrouped our study sample based on the MR findings using the difference in mean ADC value between the 1st and 2nd measurements, to analyze intraobserver reproducibility in specific situations. In the group with involved lobes > 2, there was no significant difference between the segmentation methods in the subtracted mean ADC values without reference to the structural images. However, in the group with involved lobes ≤ 2, the subtracted mean ADC value was significantly larger with the

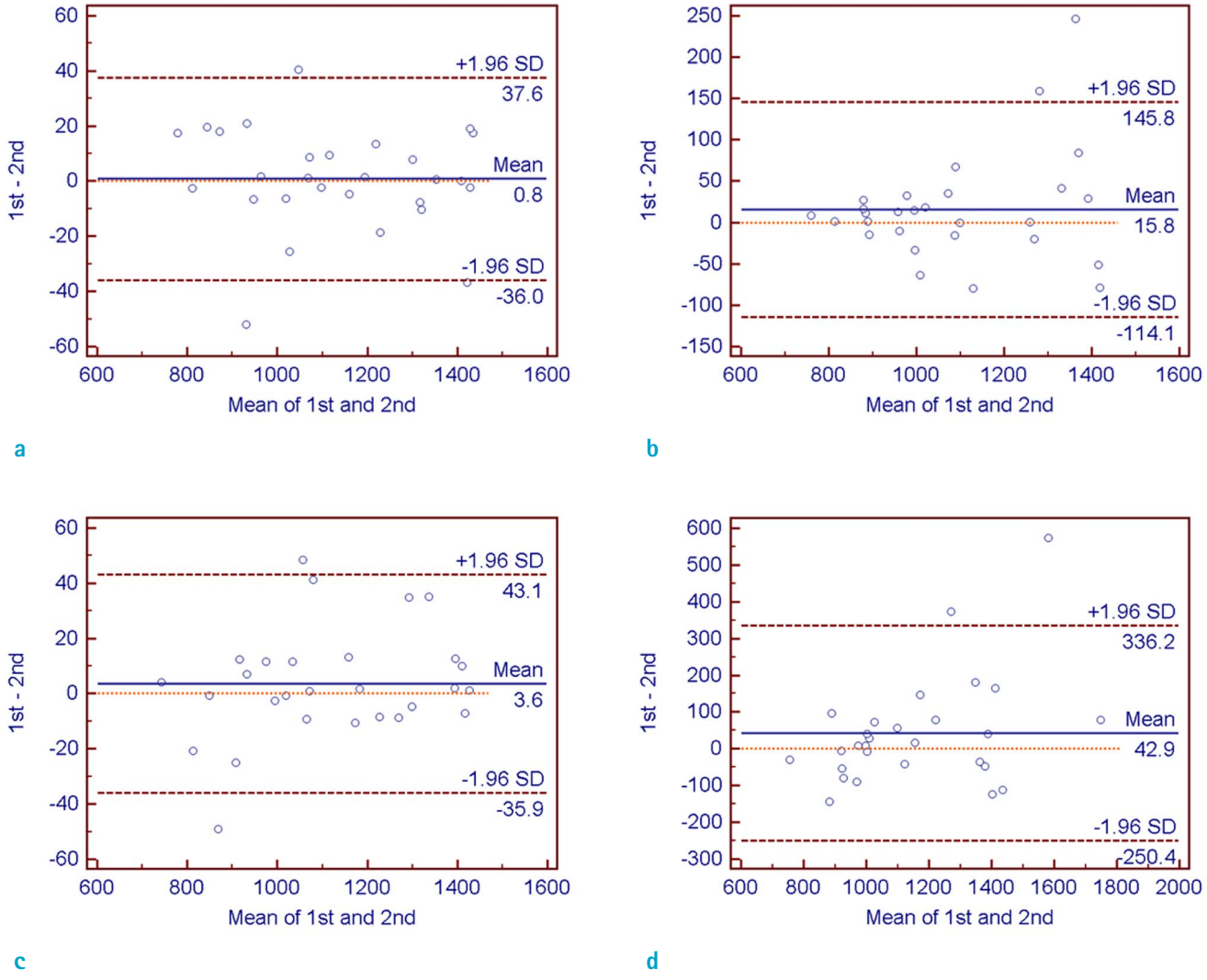

Fig. 3. Bland-Altman plots showing intraobserver reliability between the 1st and 2nd measurements in (a) the manual segmentation method with CE-T1WI for structural imaging; (b) the manual segmentation method with T2WI for structural imaging; (c) the semiautomatic segmentation method with CE-T1WI for structural imaging; and (d) the semiautomatic segmentation method with T2WI for structural imaging. CE-T1WI = contrast-enhanced T1-weighted imaging; T2WI = T2-weighted imaging

Table 3. Intraobserver Reliability in ADC Measurement

	Observer 1		Observer 2	
	Manual method			
	CE-T1WI	T2WI	CE-T1WI	T2WI
ICC*	0.996 (0.991-0.998)	0.947 (0.889-0.975)	0.986 (0.970-0.994)	0.981 (0.959-0.991)
CV†	1.15	4.35	2.28	2.33
	Semiautomatic segmentation method			
	CE-T1WI	T2WI	CE-T1WI	T2WI
ICC*	0.995 (0.990-0.998)	0.826 (0.659-0.916)	0.992 (0.982-0.996)	0.968 (0.931-0.985)
CV†	1.27	9.36	1.70	3.14

All of the numbers in brackets indicate the 95% confidence interval.

ADC = apparent diffusion coefficient; CV = coefficient of variance; CE-T1WI = contrast-enhanced T1-weighted imaging; ICC = intraclass correlation coefficient; T2WI = T2-weighted imaging

*ICC values were categorized as follows: < 0.40, poor agreement; 0.40–0.59, fair agreement; 0.60–0.74, good agreement; and ≥ 0.75, high agreement.

† Numbers are expressed as percentages.

semiautomatic segmentation method than with the manual segmentation method on T2WI-based evaluation (116.13 vs. 49.51, P = 0.0173). There were no significant differences between the methods on CE-T1WI-based evaluation on-line Table 1.

When we subclassified our study sample into 3 groups according to tumor border, there were no significant differences in the subtracted mean ADC values between the methods, either in the group with well-demarcated tumor borders or in the group with fully poorly demarcated tumor borders. However, when the tumor border was partially poorly demarcated, the subtracted mean ADC value was larger with the semiautomatic segmentation method than with the manual segmentation method, with T2WI for structural images (85.08 vs. 30.28, P = 0.0077). On CE-T1WI-based evaluation, there were no significant differences between the methods in the group with partially poorly demarcated tumor border on-line Table 2.

The proportions of cystic or solid components had no effect on the difference in the subtracted mean ADC values in any of the situations between the segmentation methods on-line Table 3. The subtracted mean ADC value was significantly higher with the T2WI-based semiautomatic segmentation method than with the T2WI-based manual segmentation method in the group with poorly demarcated inner margins of the necrotic portion (65.24 vs. 33.14, P = 0.02). In addition, there was no significant difference between the segmentation methods, regardless of structural images, when the inner margin of the necrotic portion was well demarcated on-line Table 4.

For perilesional edema, the subtracted mean ADC value was significantly larger with the semiautomatic segmentation method than with the manual segmentation method on T2WI-based evaluation (100.76 vs. 44.68, P = 0.0121). There was no significant difference in the subtracted mean ADC values with CE-T1WI for structural images on-line Table 5.

DISCUSSION

We compared interobserver and intraobserver reproducibility between the manual and semiautomatic segmentation methods using the mean ADC value, a parameter derived from DW MR imaging. The main findings of our study were as follows: (a) interobserver reliabilities were high in both segmentation methods on CE-T1WI-based evaluation only, while intraobserver reliabilities were high in both segmentation methods without reference to structural images; (b) higher reproducibility was observed with the manual segmentation method than that with the semiautomatic segmentation method on T2WI-based evaluation, even though no significant difference was found between the 2 methods on CE-T1WI-based evaluation; and (c) the reproducibility of the semiautomatic segmentation method on T2WI-based evaluation was specifically lower than that of the manual segmentation method in the groups with involved lobes ≤ 2, with partially poorly demarcated tumor borders, with poorly demarcated inner margins of the necrotic portion, and with perilesional edema.

The overall interobserver reliability was high in both segmentation methods on CE-T1WI-based evaluation only, while intraobserver reliability was high in both segmentation methods regardless of structural images, but particularly

on CE-T1WI-based evaluation. Jung et al. (19) reported that both the manual and semiautomatic segmentation methods were clinically acceptable. Moreover, they reported higher reproducibility for the semiautomatic segmentation method, as compared with the manual segmentation method on CE-T1WI-based evaluation. Our study presented slightly different results when we used mean ADC values to compare interobserver and intraobserver reliability instead of using normalized cerebral blood volume (nCBV) values. We obtained clinically acceptable reproducibility in both segmentation methods on CE-T1WI-based evaluation, but not on T2WI-based evaluation due to low interobserver reliability. Similar to the results by Jung et al. (19), our results showed higher interobserver reproducibility in the semiautomatic segmentation method than the manual segmentation method on CE-T1WI-based evaluation, even though the gap between them was negligible. Moreover, the automatic segmentation method reportedly has better reproducibility, is less time-consuming, and offers greater benefits in stratifying tumor characteristics (11, 20). However, we could not support such an idea when using the semiautomatic segmentation method. As stated above, interobserver reliability was slightly better in semiautomatic segmentation method on CE-T1WI-based evaluation than that of manual segmentation method. However, intraobserver reliability for 2 segmentation methods on CE-T1WI-based evaluation was almost the same and even more time consuming in the semiautomatic segmentation method. Thus, according to our results, there are no significant benefits in using semiautomatic segmentation method on CE-T1WI-based evaluation for obtaining histogram of the ADC values in GBM. However, CE-T1WI-based evaluation is known to be much more reproducible than T2WI-based evaluation for both manual and semiautomatic segmentation methods, showing higher interobserver and intraobsever reliability.

In addition, the manual segmentation method showed better interobserver and intraobserver reliability, as compared with that of the semiautomatic segmentation method on T2WI-based evaluation when measuring the mean ADC values. Semiautomatic segmentation using clustering analysis is performed with the expectation-maximization algorithm after manually defining the VOI. This algorithm has the limitation of being used only on T2WI to differentiate each cluster, due to the overlapping of T2 signal intensity between normal brain parenchyma and brain lesions (21). Thus, it remains challenging to develop more effective and reliable semiautomatic segmentation algorithms, particularly when using T2WI for structural images.

We further assessed intraobserver reproducibility in specific situations that were arbitrarily created according to the MR findings. Our study proved that the T2WI-based semiautomatic segmentation method was the least reliable method, particularly in groups with involved lobes ≤ 2, with partially poorly demarcated tumor borders, with poorly demarcated inner margins of the necrotic portion, and with perilesional edema, while both manual and semiautomatic segmentation methods on CE-T1WI-based evaluation are clinically acceptable for assessment of aforementioned subgroups. Relatively similar T1 and T2 relaxation parameters among pathologic brain lesions, brain edema, and fibrotic tissue are obtained on T2WI, as compared with CE-T1WI (20). It is also known that the differences in signal intensities among the clusters are not sufficient to separate tumor components, such as solid, cystic, or hemorrhagic components, perilesional edema, and perilesional tumor infiltration, automatically on T2WI (10, 22). According to the aforementioned reports, the T2WI-based semiautomatic segmentation method would be less reproducible in the groups with perilesional edema, with poorly demarcated inner margins of the necrotic portion, and with poorly demarcated outer tumor borders. Contrary to prior reports, the proportions of the tumor components, such as solid and cystic components, did not affect the difference in reproducibility between the manual and semiautomatic segmentation methods. Interestingly, the difference in the subtracted mean ADC values between the segmentation methods was only significantly different in the group with partially poorly demarcated tumor borders and not in the group with fully poorly demarcated tumor borders. When the tumor was fully poorly demarcated, it was also difficult to outline the tumor extent using the manual segmentation method possibly due to the lack of significant difference between the segmentation methods.

Our study had several limitations. First, we included only GBM in the analysis of interobserver and intraobserver reliability and excluded other types of gliomas. This criterion might have allowed for the exclusion of other types of gliomas or other brain tumors because our study was designed for the evaluation of reproducibility between 2 different segmentation techniques. However, there is a possibility of obtaining different results with other types of brain tumors because the characteristics of each brain tumor are different. Further study with larger cohorts with brain tumors might be required. Second, we used

consensus mean ADC values of GBM for assessment of various subgroups that were divided according to the MR imaging findings. Therefore, interobserver reliability in various subgroups could not be assessed in our study. Third, although we attempted to reduce recall bias during the measurement of the mean ADC values by allowing for an interval of at least 2 weeks between measurements, the possibility of recall bias could not be completely excluded, and might have critically affected the intraobserver reliability results. However, we measured mean ADC values over a period of 6 months, and the interference of recall bias in our results was extremely low in possibility. Last, our study population was not large enough for subgroup analysis. Further study with a larger study population might be required.

In conclusion, both the manual and semiautomatic segmentation methods on CE-T1WI-based evaluation were clinically acceptable in the measurement of mean ADC values, showing high interobserver and intraobserver reliabilities. It is still challenging to use the segmentation methods on T2WI-based evaluation due to low interobserver reliability, especially in semiautomatic segmentation method. T2WI-based semiautomatic segmentation method should be avoided, particularly in the subgroups of GBM with involved lobes ≤ 2, with partially poorly demarcated tumor borders, with poorly demarcated inner margins of the necrotic portion, and with perilesional edema.

Acknowledgments

This study was supported by a grant from the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (1120300), the Korea Healthcare technology R&D Projects, Ministry for Health, Welfare & Family Affairs (A112028 and HI13C0015), and the Research Center Program of IBS (Institute for Basic Science) in Korea.

REFERENCES

1. Schwartzbaum JA, Fisher JL, Aldape KD, Wrensch M. Epidemiology and molecular pathology of glioma. Nat Clin Pract Neurol 2006;2:494-503; quiz 491 p following 516
2. Law M, Young R, Babb J, Pollack E, Johnson G. Histogram analysis versus region of interest analysis of dynamic susceptibility contrast perfusion MR imaging data in the grading of cerebral gliomas. AJNR Am J Neuroradiol 2007;28:761-766
3. Sugahara T, Korogi Y, Kochi M, et al. Usefulness of diffusion-weighted MRI with echo-planar technique in the evaluation of cellularity in gliomas. J Magn Reson Imaging 1999;9:53-60
4. Kang Y, Choi SH, Kim YJ, et al. Gliomas: Histogram analysis of apparent diffusion coefficient maps with standard- or high-b-value diffusion-weighted MR imaging--correlation with tumor grade. Radiology 2011;261:882-890
5. Provenzale JM, Mukundan S, Barboriak DP. Diffusion-weighted and perfusion MR imaging for brain tumor characterization and assessment of treatment response. Radiology 2006;239:632-649
6. Murakami R, Hirai T, Sugahara T, et al. Grading astrocytic tumors by using apparent diffusion coefficient parameters: superiority of a one- versus two-parameter pilot method. Radiology 2009;251:838-845
7. Murakami R, Hirai T, Kitajima M, et al. Magnetic resonance imaging of pilocytic astrocytomas: usefulness of the minimum apparent diffusion coefficient (ADC) value for differentiation from high-grade gliomas. Acta Radiol 2008;49:462-467
8. Huo J, Okada K, Kim HJ, Pope WB, Goldin JG, Alger JR. CADrx for GBM brain tumors: predicting treatment response from changes in diffusion-weighted MRI. Algorithms 2009;2:1350-1367
9. Pope WB, Kim HJ, Huo J, et al. Recurrent glioblastoma multiforme: ADC histogram analysis predicts response to bevacizumab treatment. Radiology 2009;252:182-189
10. Fletcher-Heath LM, Hall LO, Goldgof DB, Murtagh FR. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. Artif Intell Med 2001;21:43-63
11. Clark MC, Hall LO, Goldgof DB, Velthuizen R, Murtagh FR, Silbiger MS. Automatic tumor segmentation using knowledge-based techniques. IEEE Trans Med Imaging 1998;17:187-201
12. Nie J, Xue Z, Liu T, et al. Automated brain tumor segmentation using spatial accuracy-weighted hidden Markov Random Field. Comput Med Imaging Graph 2009;33:431-441
13. Kaus MR, Warfield SK, Nabavi A, Black PM, Jolesz FA, Kikinis R. Automated segmentation of MR images of brain tumors. Radiology 2001;218:586-591
14. Corso JJ, Sharon E, Dube S, El-Saden S, Sinha U, Yuille A. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. IEEE Trans Med Imaging 2008;27:629-640
15. Bjornerud A. The ICE software package: direct co-registration of anatomical and functional datasets using DICOM image geometry information. Proc Hum Brain Mapping 2003;19:1018p
16. Sundar H, Shen D, Biros G, Xu C, Davatzikos C. Robust computation of mutual information using spatially

adaptive meshes. Med Image Comput Comput Assist Interv 2007;10:950-958
17. Pluim JP, Maintz JB, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging 2003;22:986-1004
18. Reed GF, Lynn F, Meade BD. Use of coefficient of variation in assessing variability of quantitative assays. Clin Diagn Lab Immunol 2002;9:1235-1239
19. Jung SC, Choi SH, Yeom JA, et al. Cerebral blood volume analysis in glioblastomas using dynamic susceptibility contrast-enhanced perfusion MRI: a comparison of manual and semiautomatic segmentation methods. PLoS One 2013;8:e69323
20. Emblem KE, Nedregaard B, Hald JK, Nome T, Due-Tonnessen P, Bjornerud A. Automatic glioma characterization from dynamic susceptibility contrast imaging: brain tumor segmentation using knowledge-based fuzzy clustering. J Magn Reson Imaging 2009;30:1-10
21. Prastawa M, Bullitt E, Moon N, Van Leemput K, Gerig G. Automatic brain tumor segmentation by subject specific modification of atlas priors. Acad Radiol 2003;10:1341-1348
22. Hsieh TM, Liu YM, Liao CC, Xiao F, Chiang IJ, Wong JM. Automatic segmentation of meningioma from non-contrasted brain MRI integrating fuzzy clustering and region growing. BMC Med Inform Decis Mak 2011;11:54