AJNR

Predicting Breakdown of the Blood-Brain Barrier in Multiple Sclerosis without Contrast Agents

R.T. Shinohara, J. Goldsmith, F.J. Mateen, C. Crainiceanu and D.S. Reich

ORIGINAL
RESEARCH

R.T. Shinohara
J. Goldsmith
F.J. Mateen
C. Crainiceanu
D.S. Reich

Predicting Breakdown of the Blood-Brain Barrier in Multiple Sclerosis without Contrast Agents

BACKGROUND AND PURPOSE: Disruption of the BBB in MS is associated with the development of new lesions and clinical relapses and signifies the presence of active inflammation. It is most commonly detected as enhancement on MR imaging performed with contrast agents that are costly and occasionally toxic. We investigated whether the BBB status in white matter lesions may be indirectly ascertained via examination of features on T1- and T2-weighted images obtained before the injection of a contrast agent.

MATERIALS AND METHODS: We considered 93 brain MR imaging studies on 16 patients that included T1-, T2-, and T2-weighted FLAIR images and predicted voxel wise enhancement after intravenous injection of a gadolinium chelate. We then used these voxel-level predictions to determine the presence or absence of abnormal enhancement anywhere in the brain.

RESULTS: On a voxel-by-voxel basis, enhancement can be predicted by using contrast-free measures with an AUC of 0.83 (95% CI, 0.80–0.87). At the whole-brain level, enhancement can be predicted with an AUC of 0.72 (95% CI, 0.62–0.82).

CONCLUSIONS: In many cases, breakdown of the BBB in acute MS lesions may be inferred without the need to inject an MR imaging contrast agent. The inference relies on intrinsic properties of tissue damage in acute lesions. Although contrast studies are more accurate, they may sometimes be unnecessary.

ABBREVIATIONS: AUC = area under receiver operating curve; CI = confidence interval; FA = flip angle; FSPGR = fast-spoiled gradient recalled; IQR = interquartile range; NPV = negative predictive value; ROC = receiver operating characteristic curve analysis; RRMS = relapsing-remitting multiple sclerosis; VV = voxel volume

RRMS is a potentially disabling inflammatory disease of the central nervous system, marked by symptomatic attacks that are related to the onset of demyelinating lesions in the brain and spinal cord.[1] MR imaging is crucial in the identification of new lesions, which often coincide with disturbance of the BBB but that are also associated with the presence of interstitial or intramyelinic edema.[2] MR imaging is a centerpiece of MS diagnosis. Gadolinium contrast-enhancing lesions on brain MR imaging form a critical component of the current diagnostic criteria[3] and are commonly used as outcome measures in clinical trials in the evaluation of new disease-modifying therapies.

Gadolinium contrast adds considerable expense to the evaluation of MS by MR imaging, both for the contrast agent itself and for the additional scanning time necessary to obtain the postgadolinium images. This increases the study costs in terms of both scanner time and the radiology technologist's time. Thus, although the gadolinium itself is only a fraction of the total study cost, a study ordered with contrast is significantly more expensive than a study ordered without it. For example, in the United States, Medicare will pay on average $535 (range, $444–$717) for a brain MR imaging without contrast and an average of $750[4] (range, $624–$1001) for a study with contrast. This is an average difference in price of >40%. Gadolinium may also rarely be associated with nephrogenic systemic fibrosis in patients with kidney disease.[5-7] Furthermore, gadolinium is associated with nonallergic and allergic reactions ranging in severity from mild to very rarely life-threatening.[8,9] Therefore, the necessity of the contrast agent for the identification of BBB abnormalities should be carefully evaluated on a case-by-case basis.

In this article, we show that the presence of a disrupted BBB can be predicted from unenhanced MR imaging data in patients with confirmed MS and that this information can be extracted by using standard statistical techniques. We report data from a group of patients who underwent precontrast T1- and T2-weighted imaging, followed by gadolinium-enhanced T1-weighted imaging. We use enhancement data from the postcontrast scans in the training of a voxel-level classifier, demonstrating that enhancement may be predicted by using the precontrast data. These predictions may also be used to complement standard MR imaging for the examination of BBB breakdown by a neuroradiologist.

Characteristics of patients in study at first visit	
Variable	Population (N = 16)
Male sex (No.) (%)	6 (38)
Age (years) (mean)	40 ± 9
Expanded Disability Status Scale (median) (IQR)	1.5 (1–2.5)
Disease duration (yr) (mean)	8 ± 5
Disease-modifying therapy (No.) (%)	13 (81)

Materials and Methods

The goal of this analysis was to model BBB disruption in white matter lesions by using MR imaging measures that do not involve contrast injection. We used as our criterion standard a manual segmentation of enhancing voxels by a neuroradiologist (D.S.R.) with 7 years' experience analyzing MR imaging in MS.

Study Population

Data from 93 MR imaging scans of 16 individuals with MS, obtained under an institutional review board–approved protocol, were analyzed. All participants gave informed consent. Demographic, diagnostic, and treatment information are reported in the Table. Each patient had between 2 and 10 visits and was imaged sequentially for up to 11 months. Eight patients changed disease-modifying therapy during the study period, and all patients were receiving disease-modifying treatment by the time of their last MR imaging.

Experimental Methods

Images were acquired on either a 1.5T (59 scans) or a 3T (34 scans) MR imaging scanner (Signa Excite HDxt; GE Healthcare, Milwaukee, Wisconsin) by using the body coil for transmission and an 8-channel receive coil array (Invivo, Gainesville, Florida) for signal-intensity detection. Sequence parameters varied depending on the platform. T1-weighted scans were obtained before contrast administration by using a 3D FSPGR sequence, with TR = 9 ms, TE = 3.5 ms, TI = 450 ms, FA = 13°, and nominal VV = 1–1.2 mm³. T2-weighted scans were also obtained before contrast administration by using a 2D fast spin-echo sequence (TR = 5–6 seconds, TE ∼ 120 ms, FA = 90°, VV = 1–2 mm³).

An intravenous infusion of 0.1 mmol/kg of gadopentate dimeglumine (Magnevist; Bayer HealthCare, Leverkusen, Germany) via a power injector (Solaris; Medrad, Indianola, Pennsylvania) was administered. T1-weighted scans were also obtained a median of 8 minutes (IQR, 6–15 minutes) following contrast injection by using either a 2D spin-echo sequence (TR = 600 ms, TE = 16 ms, FA = 90°, VV = 2 mm³) or the same 3D FSPGR sequence used for precontrast T1-weighted scans. In addition, T2-weighted FLAIR scans were obtained at a median of 15 minutes (IQR, 14–45 minutes) following contrast injection, again by using either a 2D fast spin-echo sequence on the 1.5T scanner (TR = 10 seconds, TE ∼ 123 ms, TI = 2250 ms, FA = 90°, VV = 3.5 mm³) or a 3D variable FA sequence on the 3T scanner (TR = 6 seconds, TE ∼ 126 ms, TI ∼ 1860 ms, VV = 1 mm³). Postcontrast T2-weighted FLAIR scans allow detection of contrast-enhancing lesions without affecting the signal intensity of nonenhancing lesions so that signal intensity within lesions on postcontrast scans is at least as great as that on precontrast scans.[10,11]

Image Postprocessing

For the image processing, Medical Image Processing Analysis and Visualization (http://mipav.cit.nih.gov) and Java Image Science Toolkit (http://nitrc.org/projects/jist) were used. All statistical calculations and

modeling were conducted by using the software environment R (Version 2.12.0; R Foundation for Statistical Computing, Vienna, Austria).

All acquired volumes were rigidly registered to the pregadolinium T1-weighted volume and then rigidly aligned to the Montreal Neurologic Institute standard template. Nonparametric intensity-nonuniformity normalization was used to address scanner-related inhomogeneity.[12] All scans were interpolated to a voxel size of 1 mm³. Skull and extracranial voxels were masked out by using a skull-stripping procedure.[13] To more completely remove the normally enhancing meninges and to focus our attention on the white matter where most enhancing voxels are located, we then removed all voxels below axial section 52 (the inferior temporal lobes) and above axial section 156 (the top of the brain). The remaining volume was eroded by 2 mm in each direction to remove much of the residual extracerebral tissue. In Fig 1, an axial section from the registered pre- and postcontrast injection T1-weighted, T2-weighted, and FLAIR images from 1 subject is shown.

Normalization

Because the T1-weighted, T2-weighted, and FLAIR images for each scan were acquired in arbitrary units of signal intensity, these images were first normalized. Individual T1-weighted images were normalized by subtracting the mean intensity of voxels in the brain and dividing by their SD; thus, normalization was carried out at the scan level, rather than the subject or population level. Because the upper tail distributions of the T2-weighted and FLAIR images were influenced by the hyperintense voxels within lesions, for these scans, the same normalization technique was applied by using trimmed means and SDs after excluding the top 5% of voxels. The 5% cutoff in this procedure was chosen empirically, but through a sensitivity analysis (omitted), we found the performance of our methods to be robust to this choice within a range of 2.5%–7.5%. Excess noise was removed by applying a Gaussian kernel–based spatial smoother over the entire brain (see On-line Appendix for details).

Criterion Standard Measure

To train our prediction model and assess its performance, we used a manual segmentation conducted by an experienced neuroradiologist.

Enhancement Prediction

We modeled the probability that each voxel enhances. To limit false-positive results, we excluded voxels with low FLAIR intensities. We then used logistic regression to model the probability that each voxel enhances in terms of the normalized T1- and T2-weighted intensities of that particular voxel. The result was a 3D map of the prediction statistic, an example of which we show in Fig 1D as an axial section from 1 subject.

The first step of our procedure was to select voxels in the top 1% of FLAIR intensity in each scan to capture candidate white matter lesion voxels; the FLAIR threshold was subjected to a sensitivity analysis, and our method is robust to changes in this value. We base our analysis on these potentially enhancing "candidate" voxels, which we found to be predominantly in white matter lesions. Note that the FLAIR data used for this study scan were acquired after contrast injection to aid in the clinical detection of enhancing lesions.[10,11] However, because both enhancing and nonenhancing lesions are hyperintense on postcontrast FLAIR scans, our threshold, which fell below the intensities of nonenhancing lesions (On-line Fig 1), ensured that both groups were included and that enhancement information in the postcontrast FLAIR data were not used for the enhancement prediction. Thus,
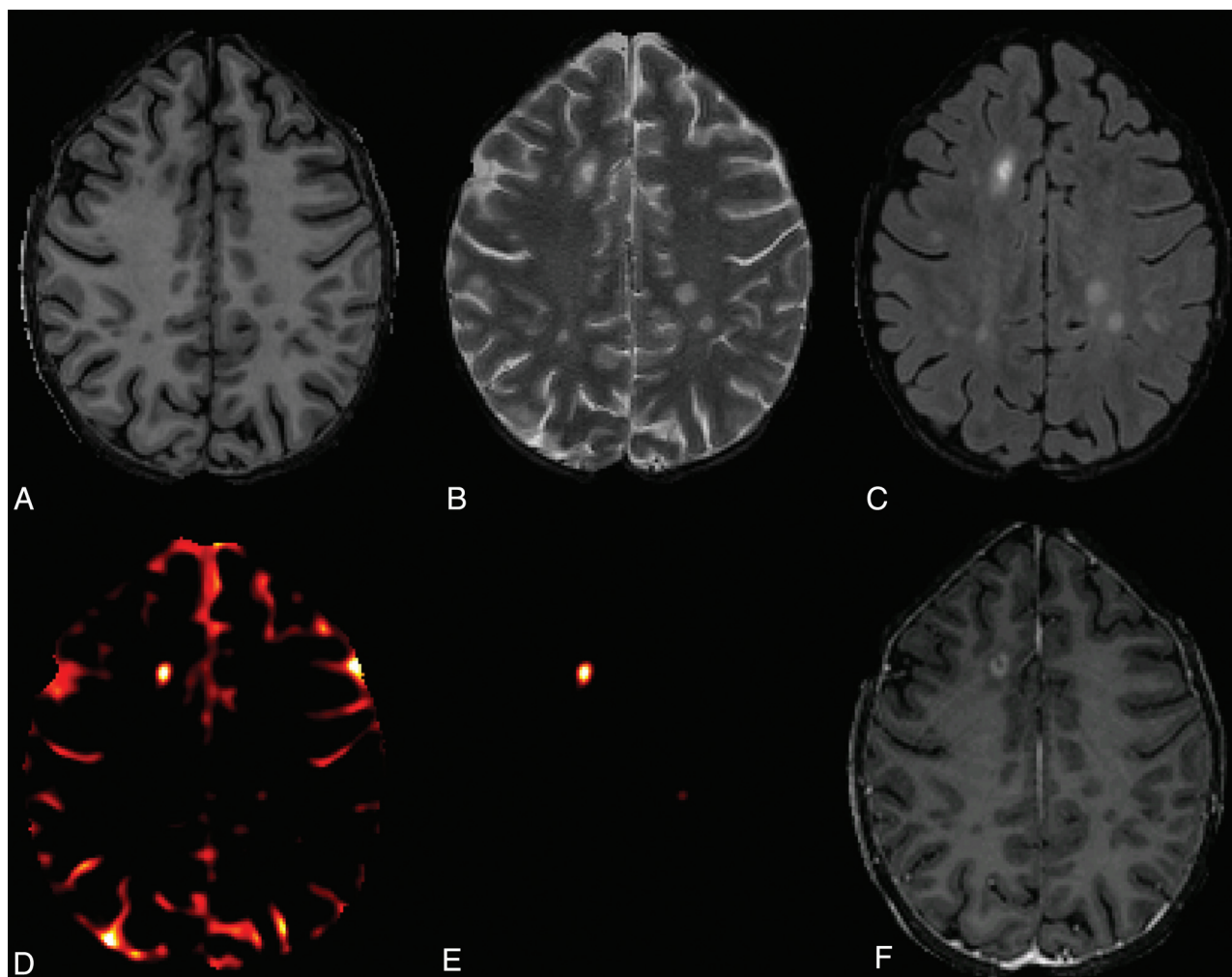
Fig 1. Axial sections from brain MR imaging (3T) of individuals with RRMS. *A*, T1-weighted precontrast. *B*, T2-weighted precontrast. *C*, T2-weighted FLAIR postcontrast. *D*, Enhancement prediction mask shown in all brain voxels for illustrative purposes (not used for analysis). *E*, Enhancement prediction mask based on model 1, shown only in candidate voxels. *F*, T1-weighted postcontrast.

voxels that met the intensity threshold criterion, regardless of whether they were enhancing or nonenhancing, were treated equally in subsequent analysis. Note that occasionally voxels within lesions may appear hypointense on FLAIR scans due to excessively long T1 relaxation times, probably reflecting tissue loss. In our experience, these voxels are found in chronic lesions, and in the data analyzed here, there were no such voxels within enhancing lesions. Exclusion of these voxels, accomplished by thresholding the FLAIR scans, therefore reduces the number of false-positives.

We modeled the probability that a voxel will enhance by using the logistic regression (model 1):

$$\text{logit Pr (voxel } v \text{ from scan } t \text{ enhances)} = \beta_0 + \beta_1 T_1(v, i)$$

$$+ \beta_2 T_2(v, i) + \beta_{12} T_1(v, i) \ T_2(v, i),$$

where $T_j(v, i)$ denotes the normalized T_j-weighted intensity of voxel v from scan i for $j = 1, 2$. This yields an estimated probability of each candidate voxel enhancing. Note that the predictors in this model were all acquired before contrast injection.

Performance Analysis

To assess the performance of these predictions, we split the data equally by subjects at random into training and validation sets. We fit model 1 on the training dataset and calculated the ROC curves and AUC for the validation dataset to assess potential overfitting. These performance measures are known to be susceptible to instability, so we nonparametrically bootstrapped (without replacement) the training and validation sets. This procedure allowed the estimation of CIs for the AUC and an average ROC curve that should be easily reproducible. We estimated the voxel-level ROC curves by using the expert manual segmentation standard measure.

To assess whether scans can be dichotomized according to the presence or absence of enhancing lesions, we built a scan-level classifier based on our voxel-level model. For this scan-level prediction, we used the maximum of the predicted probabilities from model 1 from each scan as our predictive index.

Results

The logistic regression model 1 was fit on 93 scans. Of 1.5 million candidate voxels, 7752 were enhancing according to the criterion standard. The coefficient estimates are provided in the On-line Appendix. Examination of the signs of the regression coefficients indicate that voxels with lower-than-average T1-weighted signal intensities and higher-than-average T2-weighted signal intensities are more likely to enhance.
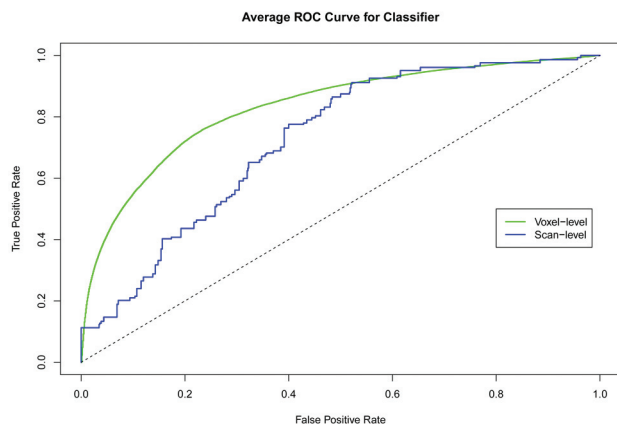
Average ROC Curve for Classifier

Fig 2. Average ROC curves for the voxel-level and scan-level predictors of enhancement in candidate white matter lesion voxels. These curves were calculated by using a bootstrap and the logistic regression model 1.

This tendency is mitigated by the positive coefficient on the interaction term, $T_1(v, i) \, T_2(v, i)$ so that voxels that are extremely hypointense on T1-weighted images and hyperintense on T2-weighted images—such as CSF—are considered less likely to be enhancing. The combination of low signal intensity on T1-weighted images and high signal intensity on T2-weighted images indicates increased water content, probably reflecting edema in acute MS lesions.[1]

The predicted values from this model are shown spatially in Fig 1*E*, where yellow indicates areas of highest predicted probability of enhancement. The voxel-level average ROC curve is shown in Fig 2 in green, and the AUC is estimated to be 0.83 (95% CI, 0.80–0.87). The scan-level results are shown in blue, and the estimated AUC is 0.72 (95% CI, 0.62–0.82).

The ROC curves in Fig 2 are similar and suggest that if given equal weight, a prediction criterion that allows better than 70% sensitivity and 70% specificity may be chosen by using an appropriate cutoff. In this study, approximately 38% of scans had at least 1 enhancing voxel, which would correspond to an NPV of 80%. Thus, for every 5 scans from this population in which our method predicts no enhancement, 1 scan will actually show enhancement if contrast is administered.

To understand the sources of error in our predictions, we investigated the scans that were incorrectly predicted to enhance (false-positives) or not to enhance (false-negatives). Most of the false-positives were due to predicted enhancement in extracerebral tissue (eg, meninges, choroid plexus, pineal gland) that was not removed by the skull-stripping or erosion procedures. In 1 subject, voxels in a single nonenhancing lesion were falsely designated as enhancing on 4 consecutive scans. Although this represents a failure of our enhancement prediction method, it indicates that some lesions are exceptional and, at the same time, demonstrates the consistency of our procedures (especially the normalization step).

Analysis of the false-negatives indicated that missed enhancing lesions were, for the most part, small enough to be blurred by the smoothing procedure—that is, on the order of a few cubic millimeters in volume. Most interesting, enhancement prediction in some lesions in 1 of the 16 subjects consistently failed on several consecutive scans. This failure suggests that the intrinsic signal-intensity properties of acute lesions may be different from 1 subject to the next, reflecting heterogeneity of the pathologic process or host response.

Discussion

Gadolinium enhancement in MR imaging is a common outcome in clinical trials for MS treatments as well as a marker of disease activity in clinical practice. Successful prediction of enhancement from standard MR imaging sequences performed without gadolinium injection would indicate that those sequences are both sensitive to and specific for tissue changes that occur in the setting of acute inflammatory lesions. Previous studies have demonstrated sensitivity of both T1- and T2-weighted scans to these changes. Our results indicate that the findings on these scans are also somewhat specific and that in the proper context and with careful interpretation, the presence or absence of enhancement may, in many cases, be inferred. For example, a lesion that is both new (ie, not present on a prior scan) and predicted to enhance may very likely truly enhance.

The modeling indicates that voxels that are moderately but not extremely hypointense on T1-weighted images and hyperintense on T2-weighted images are most likely to enhance. This finding is evident from the raw images and probably reflects the presence of increased extracellular water with only limited tissue loss in acute lesions. Water, in the form of edema, may be found in the interstitial space but also between the myelin sheets.[1,2] However, the presence of edema does not in itself imply vascular permeability, and there are many neurologic conditions in which edema is present without frank BBB opening as detected by gadolinium enhancement. Nevertheless, in the context of acute MS lesions, it may be the case that this edema directly reflects the presence of an open BBB.

From a technical point of view, our methods are computationally fast and automatic. Although training our classifier is relatively slow (by using modern computing facilities, it takes minutes to hours depending on the number of scans), the training is a 1-time procedure and the fitted model may be expressed as the 4 coefficients in model 1. The estimated coefficients were optimized for our protocol and may be useful for others, but proper training on different protocols is necessary for the performance of the classification tool. On the other hand, given a new scan, a personal computer would take only seconds to conduct the prediction based on the estimates provided in the On-line Appendix.

One way to substantially increase the power of our method would be to combine it with subtraction imaging, which can more accurately detect and identify new lesions.[14] Limiting the prediction of enhancement to lesions that are new or changed since the previous scan would likely reduce the number of false-positive results at both levels of analysis (scan and voxel). On the other hand, restricting our analysis to new or changed lesions would affect the NPV because it would change the proportion of lesions considered that are truly enhancing. Thus, the magnitude of improvement in the performance is hard to predict. Inclusion of additional unenhanced MR imaging contrasts, such as those provided by proton attenuation–weighted, diffusion-weighted, perfusion-weighted (eg, arterial spin-labeling), and magnetization-transfer imaging, which have differential sensitivity to the types of tissue damage that

occur within MS lesions, might also improve the prediction accuracy.

There are some technical limitations to the methods proposed in this article. The first is that to apply our methods optimally, removal of extracerebral tissues is required. In our implementation, we used a skull-stripping algorithm followed by an erosion procedure to remove not only the scalp, skull, and meninges but also some of the cortical mantle, which only rarely enhances in MS. However, some extracerebral tissue, such as the interhemispheric meninges, was not removed by this procedure, and this was the source of many of the voxels falsely identified as enhancing. Thus, it remains necessary to inspect the spatial location of predicted enhancements to verify that they are in the brain.

A limitation of the scan-level predictor presented in this article is that it is defined on the basis of the voxel-level classifier. This may not be optimal because spatial correlation is an important factor. Our methods address this through spatial smoothing, but more sophisticated methods may perform better. As functional and image regression techniques are developed in the statistical literature,[15] they may yield further improvement in the prediction of enhancement and measurement of BBB abnormality.

A final limitation is the use of postcontrast T2-weighted FLAIR scans to identify candidate-enhancing voxels, which we define to be those with signal intensities in the highest percentile. On the basis of our experience and on published reports,[10,11] we believe that the presence of gadolinium does not substantially alter this identification because lesions only become more hyperintense when they enhance. Unfortunately, precontrast T2-weighted FLAIR scans were not available for analysis so that direct verification of this observation on the current dataset is not possible.

Conclusions

Despite our inability to predict enhancement perfectly by using the proposed framework, we envision that our method or its successors could change clinical practice by limiting the use of contrast material in routine MR imaging of patients with established MS to cases in which there is a specific clinical question that requires contrast administration or in which the model prediction, after neuroradiologic interpretation, remains ambiguous. By reducing the need for contrast administration, automated enhancement prediction could also substantially reduce the cost of clinical trials. Whether similar methods can be applied to other diseases (such as gliomas), in which detection of enhancement is equally important, remains to be determined.

References

1. Raine CS, McFarland HF, Tourtellotte WW. *Multiple Sclerosis: Clinical and Pathogenic Basis.* London, UK: Chapman and Hall; 1997
2. Barnett MH, Prineas JW. **Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion.** *Ann Neurol* 2004;55:458–68
3. Polman CH, Reingold SC, Banwell B, et al. **Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria.** *Ann Neurol* 2011;69:292–302
4. Physician Fee Schedule. Centers for Medicare and Medicaid Services. http://www.cms.gov/apps/physician-fee-schedule/license-agreement.aspx. Accessed September 13, 2011
5. Kay J. **Nephrogenic systemic fibrosis: a gadolinium-associated fibrosing disorder in patients with renal dysfunction.** *Ann Rheum Dis* 2008;67(supple 3):iii66–69
6. Perazella MA. **Current status of gadolinium toxicity in patients with kidney disease.** *Clin J Am Soc Nephrol* 2009;4:461–69
7. Moren-Romero JA, Sequira S, Mascaro JMJ, et al. **Nephrogenic systemic fibrosis: a case series suggesting gadolinium as a possible aetiological factor.** *Br J Dermatol* 2007;157:783–87
8. Dillman JR, Ellis JH, Cohan RH, et al. **Frequency and severity of acute allergic-like reactions to gadolinium-containing i.v. contrast media in children and adults.** *AJR Am J Roentgenol* 2007;189:1533–38
9. Hunt CH, Hartman RP, Hesley GK. **Frequency and severity of adverse effects of iodinated and gadolinium contrast materials: retrospective review of 456,930 doses.** *AJR Am J Roentgenol* 2009;193:1124–27
10. Kataoka H, Taoka T, Ueno S. **Early contrast-enhanced magnetic resonance imaging with fluid-attenuated inversion recovery in multiple sclerosis.** *J Neuroimaging* 2009;19:246–49
11. Bagnato F. **Uncovering and characterizing multiple sclerosis lesions: the aid of fluid-attenuated inversion recovery images in the presence of gadolinium contrast agent.** *J Neuroimaging* 2009;19:201–04
12. Sled JG, Zijdenbox AP, Evans AC. **A nonparametric method for automatic correction of intensity nonuniformity in MRI data.** *IEEE Trans Med Imaging* 1998;17:87–97
13. Carass A, Wheeler MB, Cuzzocreo J, et al. **A joint registration and segmentation approach to skull stripping.** In: *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Chicago, Illinois; March 30-April 2, 2011:656–59
14. Moraal B, Wattjes MP, Geurts JJ, et al. **Improved detection of active multiple sclerosis lesions: 3D subtraction imaging.** *Radiology* 2010;255:154–63
15. Di CZ, Crainiceanu CM, Caffo BS, et al. **Multilevel functional principal component analysis.** *Ann Appl Stat* 2009;3:458–88