

















An Automated Statistical Technique for Counting Distinct Multiple Sclerosis Lesions

 J.D. Dworkin,  K.A. Linn,  I. Oguz,  G.M. Fleishman,  R. Bakshi,  G. Nair,  P.A. Calabresi,  R.G. Henry,  J. Oh,  N. Papinutto,  D. Pelletier,  W. Rooney,  W. Stern,  N.L. Sicotte,  D.S. Reich,  R.T. Shinohara, and the North American Imaging in Multiple Sclerosis Cooperative



ABSTRACT

BACKGROUND AND PURPOSE: Lesion load is a common biomarker in multiple sclerosis, yet it has historically shown modest association with clinical outcome. Lesion count, which encapsulates the natural history of lesion formation and is thought to provide complementary information, is difficult to assess in patients with confluent (ie, spatially overlapping) lesions. We introduce a statistical technique for cross-sectionally counting pathologically distinct lesions.

MATERIALS AND METHODS: MR imaging was used to assess the probability of a lesion at each location. The texture of this map was quantified using a novel technique, and clusters resembling the center of a lesion were counted. Validity compared with a criterion standard count was demonstrated in 60 subjects observed longitudinally, and reliability was determined using 14 scans of a clinically stable subject acquired at 7 sites.

RESULTS: The proposed count and the criterion standard count were highly correlated ($r = 0.97, P < .001$) and not significantly different ($t_{59} = -.83, P = .41$), and the variability of the proposed count across repeat scans was equivalent to that of lesion load. After accounting for lesion load and age, lesion count was negatively associated ($t_{58} = -2.73, P < .01$) with the Expanded Disability Status Scale. Average lesion size had a higher association with the Expanded Disability Status Scale ($r = 0.35, P < .01$) than lesion load ($r = 0.10, P = .44$) or lesion count ($r = -.12, P = .36$) alone.

CONCLUSIONS: This study introduces a novel technique for counting pathologically distinct lesions using cross-sectional data and demonstrates its ability to recover obscured longitudinal information. The proposed count allows more accurate estimation of lesion size, which correlated more closely with disability scores than either lesion load or lesion count alone.

ABBREVIATIONS: C_c = count based on the standard connected-components technique; C_G = criterion standard count; C_p = count based on the technique proposed in this study; CV = coefficient of variation; EDSS = Expanded Disability Status Scale; $EDSS_{avg}$ = average of the EDSS scores over all visits for each subject in the National Institute of Neurological Disorders and Stroke longitudinal study; NAIMS = North American Imaging in Multiple Sclerosis; OASIS = Automated Statistical Inference for Segmentation

Multiple sclerosis is a neuroinflammatory disorder characterized by demyelinating lesions that occur in the central nervous system. MR imaging is the most commonly used method to observe these lesions, especially in the white matter of the brain.¹

The presence of new lesions on MR imaging is often considered an important clinical marker of disease activity, yet MR imaging–based measures of disease severity have been elusive.² The total lesion burden in the white matter or “lesion load”—measured as volume or volume fraction of brain size—is often used in the

Received October 30, 2017; accepted after revision December 10.


From the Departments of Biostatistics, Epidemiology, and Informatics (J.D.D., K.A.L., R.T.S.), and Radiology (I.O., G.M.F.), Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; Laboratory for Neuroimaging Research (R.B.), Partners Multiple Sclerosis Center, Ann Romney Center for Neurologic Diseases, and Departments of Neurology (R.B.) and Radiology (R.B.), Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts; Translational Neuroradiology Section (G.N., D.S.R.), National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland; Department of Neurology (P.A.C., J.O., D.S.R.), the Johns Hopkins University School of Medicine, Baltimore, Maryland; Department of Neurology (R.G.H., N.P., W.S.), University of California, San Francisco, San Francisco, California; Keenan Research Centre for Biomedical Science (J.O.), St. Michael’s Hospital, University of Toronto, Toronto, Ontario, Canada; Department of Neurology (D.P.), Keck School of Medicine, University of Southern California, Los Angeles, California; Advanced Imaging Research Center (W.R.), Oregon Health & Science University,

Portland, Oregon; and Department of Neurology (N.L.S.), Cedars-Sinai Medical Center, Los Angeles, California. A complete list of the NAIMS participants is provided in the acknowledgment section.

This work was supported in part by the National Institutes of Health grants R01 NS085211, R21 NS093349, R01 NS094456, and R01 NS060910 from the National Institute of Neurological Disorders and Stroke. The study was also supported by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke and the Race to Erase MS Foundation.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Please address correspondence to Jordan D. Dworkin, 218 Blockley Hall, 423 Guardian Dr, Philadelphia, PA 19104; e-mail: jdwor@penmedicine.upenn.edu; @jddwor

 Indicates open access to non-subscribers at www.ajnr.org

<http://dx.doi.org/10.3174/ajnr.A5556>

study of MS, typically as a measure of disease severity³ and as a clinical trial outcome.⁴ However, lesion load has consistently shown a surprisingly weak association with clinical measures of disease severity, calling into question its usefulness as a surrogate and reinforcing the need for further development of MR imaging outcomes for MS.^{2,5}

In past years, several clinical studies have discussed the number of lesions in a patient's brain as a possible outcome of interest.⁶⁻⁸ In these studies, baseline lesion count has been shown to be correlated with the Expanded Disability Status Scale (EDSS) and changes in lesion count have been shown to be correlated with changes in the EDSS. However, obtaining an accurate count of biologically distinct lesions in the brain can be costly and logistically challenging, typically requiring expert review or automated analysis of scans obtained at frequent follow-up visits. This process is especially difficult in patients with a high lesion load and many confluent lesions.⁹

Confluent lesions commonly occur when pathologically distinct lesions (ie, lesions that arise due to spatially separate sources of structural damage in the brain, usually separated in time) occur near each other and create a larger connected region of lesion tissue. Depending on the level of lesion burden, confluent lesions can range from 2 overlapping lesions with a single connecting edge to dozens of connected lesions spanning large stretches of white matter. The existence of such confluent tissue can make it difficult or impossible to estimate the number of distinct lesions in the brain at any given visit. A patient must be scanned regularly, with temporality of appearance serving to separate spatially confluent lesions, to obtain accurate lesion counts. However, MR imaging scans are extremely costly, which can make regular follow-up visits infeasible. Additionally, in patients with a great deal of disease activity, even monthly or bimonthly scans can produce multiple new lesions that overlap in space.^{10,11} These considerations render lesion counts unavailable or inaccurate in most clinical settings in which patients are typically scanned yearly or twice yearly.

To address this issue, the current study introduces a statistical analysis technique for obtaining valid and reliable estimates of lesion counts from a single cross-sectional MR imaging study. This fully automated method uses cutting-edge statistical models for segmenting lesion tissue and well-demonstrated mathematic methods for quantifying texture to obtain the number and location of temporally distinct white matter lesions. Additionally, this study provides evidence that the derived lesion counts are associated with clinical measures of disease severity, independent of total lesion volume.

MATERIALS AND METHODS

Proposed Lesion-Count Algorithm

To obtain the lesion-count estimate in a given subject, we performed the following steps: First, a map of lesion probability at each voxel in the brain was obtained using preprocessed and coregistered MR imaging volumes from a single visit. Depending on the automated segmentation method used, a combination of T1-weighted, fluid-attenuated inversion recovery, T2-weighted, and proton density volumes would be required for probability estimation. A threshold was then applied to the prob-

ability map to create a binary mask of regions considered lesion tissue.

With the probability map, the texture of the lesion tissue was quantified to find regions that exhibited the properties expected of the center of a single lesion. Texture was quantified using the eigenvalues of the Hessian matrix. The Hessian matrix was calculated for the intensity of the lesion probability map at every voxel in the lesion mask, with a gradient window of 1 voxel in each direction. In the context of a 3D image, the Hessian matrix describes the second-order variation in image intensity in the local neighborhood around a voxel. When applied to a lesion probability map, the eigenvalues of the Hessian matrix at each voxel represent the 3 primary directions of change in lesion probability at that voxel.

Thus, voxels in the center of a lesion would be expected to have a negative eigenvalue, implying a decrease in probability, in all directions. This follows from the commonly accepted pathology of MS lesions, in which initial damage to a vein causes residual inflammation to spread outward from the vein in a relatively ovoid fashion, with less damage occurring around the periphery of the visible lesion.¹² Therefore, voxels are eliminated if any of the 3 eigenvalues are positive; the elimination indicates that the voxel is less likely to be a lesion than its surroundings in at least 1 direction. The remaining voxels with 3 negative eigenvalues are clustered by location, and connected clusters (operationalized as the centers of distinct lesions) are counted. Figure 1 provides an example of this technique.

Data and Preprocessing

Validation and Clinical-Radiologic Association. Sixty subjects diagnosed with MS were scanned between 2000 and 2008 monthly for ≤ 5.5 years (mean, 2.2 ± 1.2 years) as part of a natural history study at the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland. The subjects ranged from 18 to 60 years of age, with a mean age of 38 ± 9 years. Of the 60 subjects, 38 were women and 22 were men. Most subjects ($n = 44$) were diagnosed with relapsing-remitting MS; 13 had secondary-progressive MS; 1 had primary-progressive MS; and 2 were unspecified. Each subject was either untreated or treated with a variety of disease-modifying therapies during the observation period, including both FDA-approved (various preparations of interferon- β) and experimental therapies.

Details of the image acquisition and preprocessing have been previously published¹³ and are briefly summarized in this section. Whole-brain 2D FLAIR, proton density, T2, and 3D T1-weighted volumes were acquired on a 1.5T MR imaging scanner (Signa Excite HDxt; GE Healthcare, Milwaukee, Wisconsin). The 2D FLAIR, proton density, and T2 volumes were acquired using fast-spin-echo sequences, and the 3D T1 volume was acquired using a gradient-echo sequence. All scanning parameters were clinically optimized for each acquired image. Each subject was scanned over multiple visits, and subjects' images at each visit were rigidly coregistered longitudinally and across sequences to a template space.¹⁴

All images were N4 bias-corrected, and FLAIR, T2, and proton density volumes for each subject were interpolated and rigidly coregistered to the T1 volume in isotropic 1-mm³ space.¹⁵ Extra-

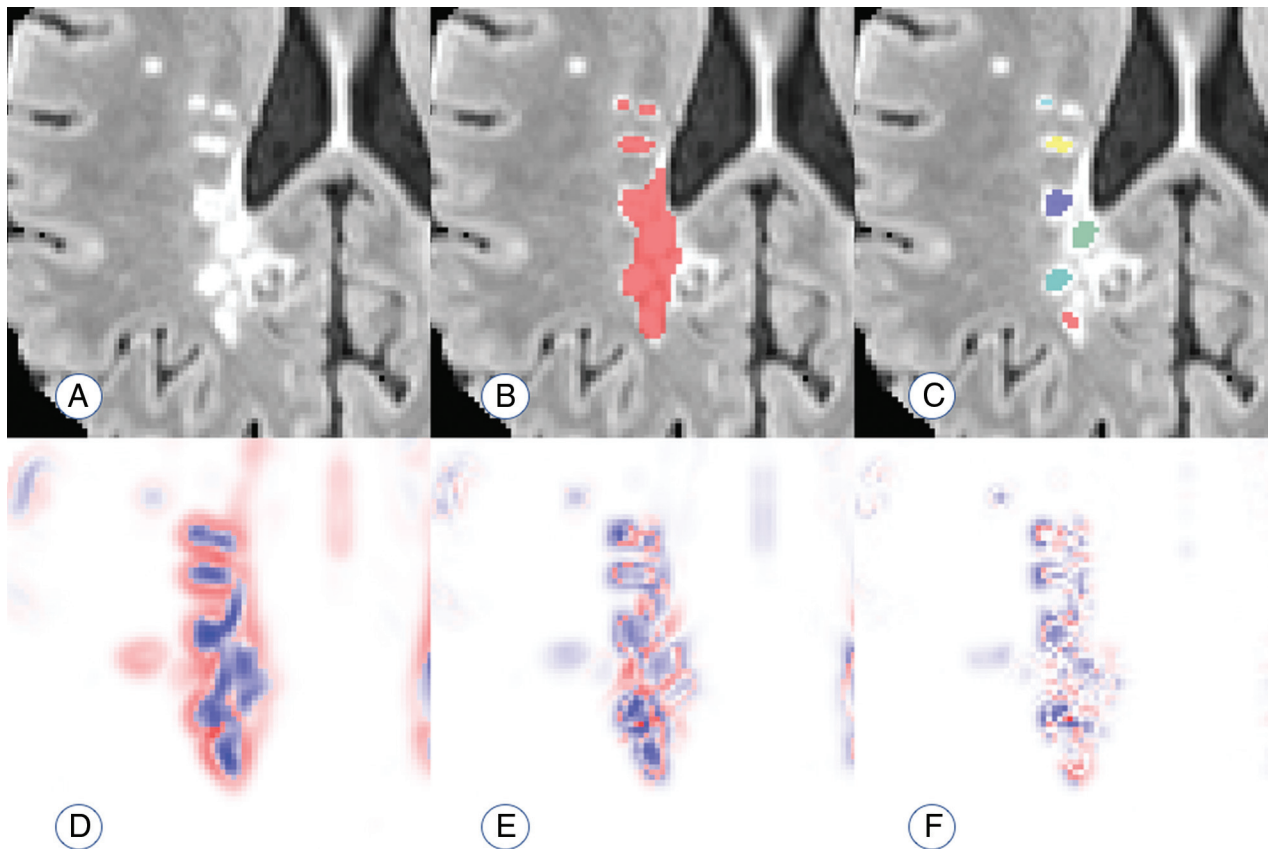


FIG 1. Visualization of the performance of the proposed count. *A*, An example of a region with confluent lesion tissue. *B*, Connected-components method finds 1 confluent lesion in the *highlighted region*. *C*, The proposed method finds 6 distinct lesion centers in the *highlighted region* on the visualized axial slice. *D–F*, Maps of the 3 Hessian eigenvalues used to quantify lesion texture and find distinct lesion centers (red represents positive eigenvalues; blue represents negative eigenvalues).

cerebral voxels were removed using the T1 volume via a skull-stripping procedure,¹⁶ and intensity normalization¹⁷ of the volumes based on *z* scoring was applied. Studies were manually quality-controlled by a researcher with >5 years' experience with structural MR imaging, and studies with analysis-limiting motion or other artifacts were removed. Following preprocessing and quality control, automatic lesion segmentation was performed on coregistered T1, T2, FLAIR, and proton density volumes using the Automated Statistical Inference for Segmentation (OASIS) model¹⁸ to produce a lesion probability map for each subject. A conservative threshold of 30% was applied to the probability maps to create binary lesion masks.

Reliability. To test reliability, also referred to as repeatability, we analyzed data from a 45-year-old man diagnosed with clinically stable relapsing-remitting MS. This patient was imaged at 7 sites in the United States as part of a pilot study for the North American Imaging in Multiple Sclerosis (NAIMS) Cooperative. He was characterized as having mild-to-moderate physical disability, which was stable between the first and last visits, and had no clinical relapses or radiologic changes during the study.¹⁹

Details of the image acquisition have been previously published¹⁹ and are briefly summarized in this section. Whole-brain 3D high-resolution FLAIR, T2, and T1-weighted volumes were acquired on 7 3T MR imaging scanners across the United States (4 Skyra, 2 Tim Trios, 1 Verio; Siemens, Erlangen, Germany). A

standardized high-resolution scanning protocol was developed through a consensus agreement in the NAIMS Cooperative and was used to the extent possible (allowing for different scanner types and software versions) for each scan. The participant was scanned twice on the same day at each site and was removed and repositioned between scan and rescan.

All images were N4 bias-corrected, and the subject's images at each scan were rigidly coregistered across sequences to the T1 volume in isotropic 1-mm³ space.¹⁵ Extracerebral voxels were removed using the T1 volume via a skull-stripping procedure,²⁰ and intensity normalization¹⁷ of the volumes based on *z* scoring was applied. Following preprocessing, automatic lesion segmentation was performed on coregistered T1, T2, and FLAIR volumes using an extension of the OASIS model²¹ to produce a lesion-probability map for each scan session. A conservative threshold of 30% was applied to the probability maps to create binary lesion masks.

Statistical Analysis

Validation. Using the longitudinal nature of the data, we developed a criterion standard count of lesions that appeared during the study for validation. A state-of-the-art technique for segmenting new lesions since a previous visit²² was applied at each visit after baseline, resulting in the number and location of new lesions at each visit for every patient. For the criterion standard count, segmented regions containing lesions separated in space or time

were considered distinct. For example, if a large contiguous region at the end of a study consisted of 1 lesion that appeared at the sixth visit and 1 lesion that appeared at the eighth visit, these would be considered 2 lesions in the criterion standard count.

The criterion standard count, henceforth referred to as C_G , was compared with 2 counts obtained cross-sectionally at the final observation for each patient. The first, C_p , is the count based on the technique proposed in this study. C_p was obtained by applying the algorithm described in the “Proposed Lesion-Count Algorithm” section to the images obtained at each patient’s final visit, then restricting the count to the number of lesion centers contained in the lesion voxels determined to have appeared during the study. Most important, this restriction means that C_p represents a subset of the total number of lesions in a subject’s scan and is distinct from the full lesion count later described in the context of the clinical-radiologic analysis. This limitation was implemented to make direct comparison between C_p and C_G possible because a criterion standard count can only be obtained for lesions that appeared during the study.

The second cross-sectional count, C_C , refers to a count based on the standard connected-components technique. C_C was obtained by performing lesion segmentation on the images obtained at each patient’s final visit, thresholding at a probability of 30%, and labeling lesions as distinct if they were separated in space. C_C was then restricted to the number of unique lesion labels contained in the lesion voxels known to have appeared during the study, to facilitate comparison with C_p and C_G .

Comparison among C_G , C_C , and C_p occurred in 2 ways: First, to compare the linear correspondence between the criterion standard and the different counting techniques, we compared the correlation between C_G and C_p with that of C_G and C_C . Then, to determine whether the counts themselves differed meaningfully from the criterion standard, paired t tests were run for C_G and C_p , as well as C_G and C_C .

Reliability. Determination of the reliability of the proposed counting method was based on the coefficient of variation (CV) of the counts obtained from the 14 repeat scans. Because the typical connected-components technique for counting automatically or manually segmented lesions yields a stable-but-invalid estimate of the true count, there is no current criterion standard CV for a lesion count. Thus, the CV of the proposed count was compared with a commonly used outcome measure for MS: total cerebral lesion volume (“lesion load”).

This comparison took place in 2 contexts. The first represented a fully automated version of the proposed count, in which variation may arise from false-negatives in the segmentation mask, false-positives in the segmentation mask, thresholding of the segmentation mask, and changes in the Hessian structure of the segmentation mask. This coefficient was compared with the CV of the automated lesion load, as determined by the segmentation method.

The second context represented a manually supplemented version of the count, in which a mask of lesion tissue was provided by an expert rater¹⁹ and the count was obtained using the segmentation probability map within the manual lesion tissue mask. In this case, variation in the count arises solely due to changes in the Hessian structure of the segmentation mask and changes in the

manual segmentation. This coefficient was compared with the CV of the manually obtained lesion load.

Clinical-Radiologic Association. Because the Expanded Disability Status Scale score is known to be noisy, a more stable measure of neurologic disability was created by averaging the EDSS scores over all visits for each subject in the National Institute of Neurological Disorders and Stroke longitudinal study,¹³ hereby referred to as the $EDSS_{avg}$. One subject had no EDSS information across all follow-ups and was excluded from this analysis. Using the OASIS lesion probability maps,¹⁸ we obtained the lesion load at the final visit for each subject using a probability threshold of 30%. Then, using the lesion-count technique described in the “Proposed Lesion-Count Algorithm” section, we obtained a full count of white matter lesions at the final visit for each subject. Most important, the counts obtained for the clinical-radiologic analysis are distinct from the C_p measure described in the “Validation” section because these counts represent the application of the proposed method to the entire brain, while C_p represents the application of the proposed method to only lesion tissue that appeared during the longitudinal study.

To determine the clinical relevance of the proposed lesion count independent of other potentially confounding variables, we created a linear regression model for $EDSS_{avg}$, with age, lesion load, and lesion count as predictors. The added statistical contribution of the lesion count was quantified using a Wald test, which is inferentially identical to a likelihood ratio test in this context, and its added clinical contribution was quantified by the increase in the adjusted R^2 of the model. In this context, R^2 gives the amount of variation in $EDSS_{avg}$ explained by the model. Additionally, the Pearson correlations with $EDSS_{avg}$ were calculated for lesion load and lesion count and a new variable we refer to as “average lesion size” (defined as lesion load divided by lesion count).

RESULTS

Validation

The temporally informed criterion standard count of new lesions appearing during the study, C_G , ranged from 0 to 75 among the 60 subjects, with a median of 4 (interquartile range, 1–12). The connected-components count, C_C , ranged from 0 to 14 with a median of 2 (interquartile range, 1–5). The proposed count, C_p , ranged from 0 to 60 with a median of 4 (interquartile range, 1–15). Figure 2 provides an example of these counting techniques.

The correlation between C_p and C_G was 0.97, compared with the correlation of 0.67 between C_C and C_G . Figure 3 shows the scatterplots for the 2 linear associations, along with the line demonstrating a 1-to-1 relationship. The paired t test comparing C_C and C_G yielded a highly significant result ($t_{59} = 4.19$, $P < .001$), with C_G being 6.9 lesions larger than C_C on average (95% CI, 3.6–10.2). The paired t test comparing C_p and C_G did not find a significant difference between the counts ($t_{59} = -.83$, $P = .41$), with C_p being 0.4 lesions larger than C_G on average (95% CI, -1.3 –0.5).

Reliability

For the fully automated count, the coefficient of variation was 0.19, compared with a CV of 0.22 for the automated lesion load.

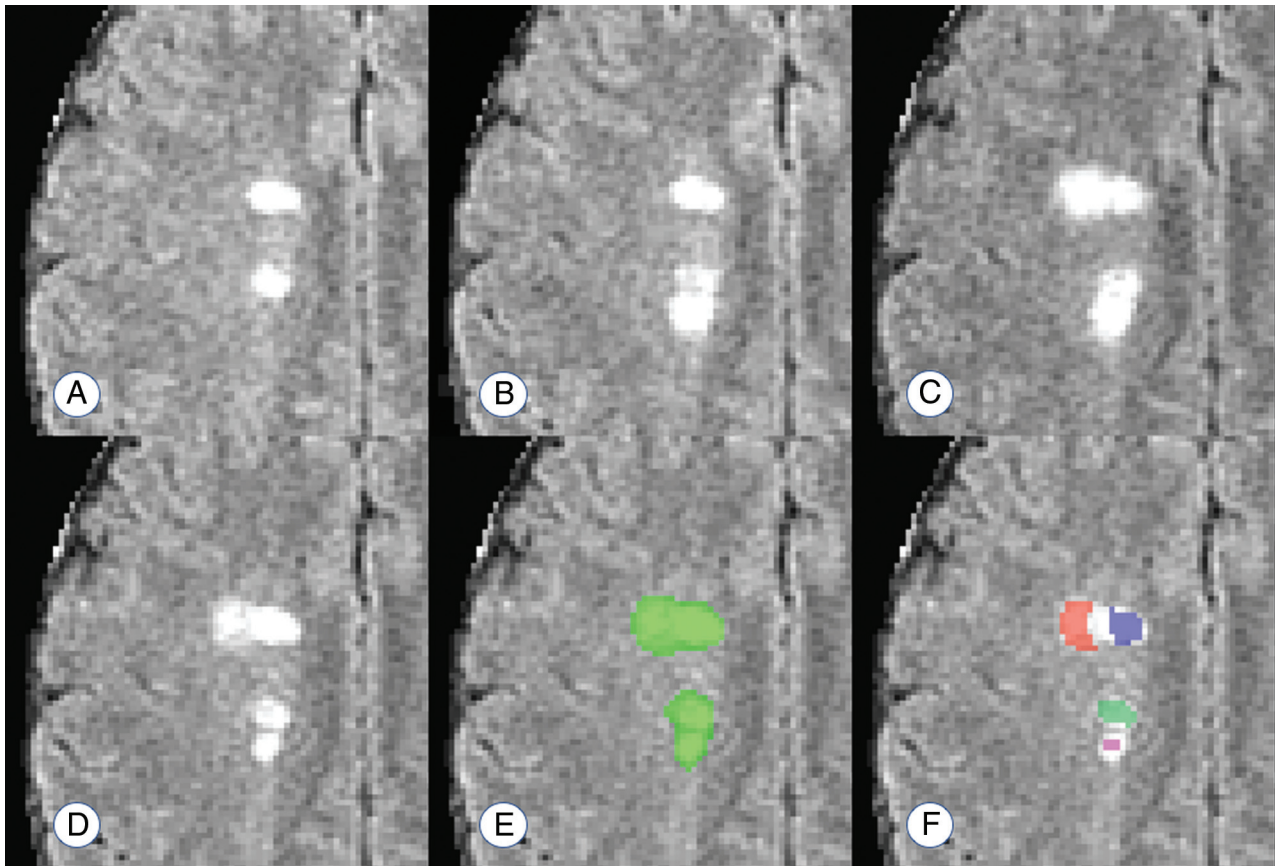


FIG 2. Example of the lesion counts in a region with 4 apparently distinct lesions, 2 of which develop with observable temporal separation. A–D, Development of 2 new and temporally distinct lesions. E and F, The performance of a connected-components count and the proposed count, respectively. The connected-components method finds 1 confluent lesion in the visualized space (connected in an adjacent plane), and the proposed method finds 4 distinct lesion centers. Days from scan in A: 28 days (B); 91 days (C); 252 days (D–F).

Using the manual segmentation as a mask, we reduced the CV for the lesion count to 0.12, compared with a CV of 0.10 for the manual lesion load. In 1 case, automated lesion segmentation was discovered to have failed; this failure created a probability map with a drastically different Hessian structure and large regions of false-positive segmentation. With this scan removed, the CV of the fully automated lesion count remained at 0.19 and the CV of the manual segmentation-based lesion count dropped to <0.06 , suggesting that the proposed count has equivalent or lower variability than the current clinical standard of lesion load.

Clinical-Radiologic Association

If we accounted for lesion load and age, the proposed lesion count was negatively associated with $EDSS_{avg}$ ($t_{58} = -2.73, P < .01$); this finding suggests that for a given lesion load and age, a higher count is associated with lower disease severity. The inclusion of lesion count in the model explains an additional 10% of the variance in $EDSS_{avg}$ compared with a model with only age and lesion load, providing support to the hypothesis that the proposed count contains disease information independent of other commonly used measures.

The Pearson correlation between lesion load and $EDSS_{avg}$ was small and did not reach significance ($r = 0.10, P = .44$); the same was true of the correlation between lesion count and $EDSS_{avg}$ ($r = -.12, P = .36$). However, average lesion size was significantly

correlated with $EDSS_{avg}$ ($r = 0.35, P < .01$); this correlation indicated that larger lesions were associated with higher disability.

DISCUSSION

In this article, we introduce a novel technique for obtaining cross-sectional counts of pathologically distinct lesions and demonstrate it to be a valid, reliable, and clinically meaningful biomarker for MS disease status. Using information contained in the Hessian structure of lesion probability maps produced by automated segmentation methods, this technique counts distinct lesions by identifying regions that resemble the physiologic traits of distinct lesion centers.

The validity of this measure was established by comparing counts obtained at a single time point with criterion standard counts that incorporated temporal information on lesion development. The proposed count had a correlation of 0.97 with the criterion standard count, indicating the very strong validity of this measure. A count obtained using the connected-components method had only a 0.67 correlation with the criterion standard and appeared to strongly underestimate the number of lesions in individuals who developed >1 or 2 lesions per year during the study. This underestimation manifested in a highly significant difference between the connected-components counts and the criterion standard counts in a paired t test, whereas no difference was found between the

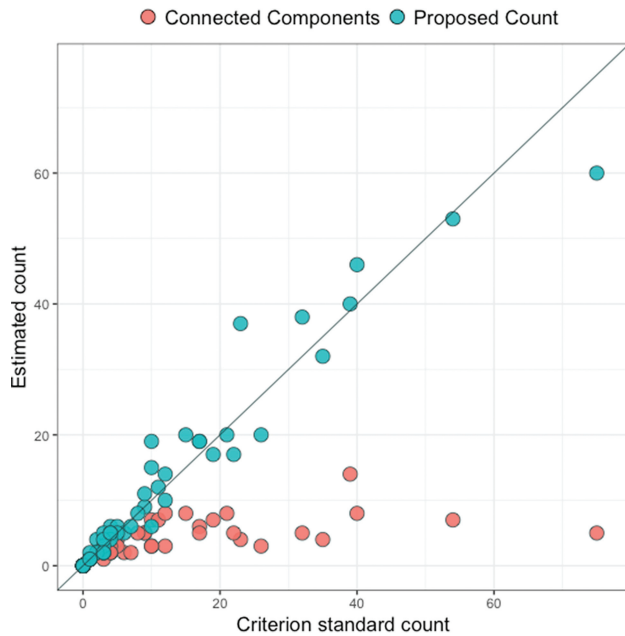


FIG 3. Scatterplot for the comparison between the criterion standard count and the connected-components count and the comparison between the criterion standard count and the proposed count. The diagonal line represents a 1-to-1 relationship, red points represent the connected-components count, and blue points represent the proposed count.

proposed counts and the criterion standard counts. These findings demonstrate that the proposed technique yields a count consistent with the natural history of lesion formation.

Reliability was considered using a rich set of data from the NAIMS Cooperative.^{19,23} In the NAIMS pilot study,²⁴ a clinically and radiologically stable subject was scanned 2 times at each of 7 different sites across the United States. The lesion count was obtained for all 14 scans of this subject, and the coefficient of variation of the counts was compared with that of lesion load in 2 contexts, to judge the reliability of the proposed measure. In the fully automated comparison, lesion count had a slightly lower CV than lesion load. This finding indicates that across repeat scans of the same brain, automated lesion count is a less variable measure than automated lesion load. In the manually supplemented comparison, lesion count had a slightly higher CV than lesion load, implying that manually obtained lesion load is a slightly less variable measure than semiautomated lesion count. On inspection, there appeared to be 1 scan in which automated lesion segmentation failed, producing an abnormal Hessian structure within the manually segmented lesion mask. With this scan removed, the CV of the semiautomated lesion count dropped to slightly more than half that of the manual lesion load. This result suggests that when automated lesion-segmentation methods perform as expected, the semiautomated lesion count is appreciably more reliable than the manual lesion load, a widely used measure of disease severity.

Clinically, the lesion-count measure appears to be a potentially important addition to commonly used radiologic biomarkers for MS. In a model accounting for lesion load and age, lesion count was highly significantly associated with EDSS. Most interesting, this association was negative, indicating that for subjects

who have similar lesion loads, better outcomes are associated with more (and smaller) lesions rather than fewer (and larger) lesions. This finding lends support to the idea that neither the number of lesions nor the amount of tissue damage alone captures all relevant clinical information and instead suggests that they should be considered together. One way to conceptualize the combination of these metrics is average lesion size, which taps into the degree to which the brain can halt the growth of lesions and encourage lesional recovery^{13,25,26} after incidence.

To investigate this concept more directly, we created a measure of average lesion size by dividing lesion load by lesion count. Pearson correlations with EDSS were then compared for the 3 biomarkers: lesion load, lesion count, and average lesion size. These findings provided further support for the combined importance of lesion load and lesion count, with both showing small and nonsignificant associations with EDSS. However, average lesion size showed a significant positive association with EDSS, consistent with the notion that the ability of the brain to slow or stop lesion growth is clinically relevant. These findings point to the importance of considering lesion count in MS research and provide further evidence of the validity of the proposed counting technique.

A limitation of the current study is the possibility of alternate explanations of confluence that are not accounted for in the design of the proposed count. It has been hypothesized that confluent lesions may occasionally occur as a result of the growth of older lesions or the expansion of pathologic processes. Future research should consider the degree to which this technique does or does not characterize these types of confluence as pathologically distinct lesions. Additionally, the current analyses do not account for the possibility of vascular comorbidity, which is a common and notable occurrence in patients with MS. Future work should investigate the performance of this algorithm in the presence of vascular lesions.

The lesion-count method presented in this article has several appealing features, including its low computational burden and its easy and flexible implementation. Computationally, the counting algorithm takes less than a minute to run once probability maps are obtained. The speed of the full technique varies depending on the lesion-segmentation method used but took approximately 25 minutes per subject as presented in this study. In terms of implementation, this method can be quickly and easily coded in any program capable of calculating the Hessian structure of a 3D image, a feature included in most image-processing packages. It can also be used with any lesion-segmentation method that yields a probability map; thus, it may be added to almost any pipeline regardless of the preferred segmentation algorithm.

CONCLUSIONS

This article introduces a novel and reliable fully automated method for counting pathologically distinct lesions using images obtained at a single time point, allowing an accurate reconstruction of the natural history of lesion formation without longitudinal data. Lesion count was found to be significantly associated with EDSS, independent of potential confounders such as lesion load and age, and the results suggest that individuals with more

small lesions may have better clinical outcomes than those with fewer large lesions. This study also demonstrates the importance of obtaining both lesion count and lesion load by using them to construct a new MS biomarker, average lesion size, and showing that average lesion size has a significantly larger association with EDSS than both lesion load and lesion count. With further study, this technique and the findings it produces could set the stage for new lesion-level considerations in research and treatment of MS.

ACKNOWLEDGMENTS

The authors would like to thank Blake Dewey, the Neuroimmunology Branch clinical group, and the technicians at the National Institutes of Health who were instrumental in helping to collect and process the validation and clinical-radiologic data. Additionally, the following is a full list of individuals who contributed to the collection and processing of the NAIMS data used in the reliability analysis described in this article: Brigham and Women's Hospital, Harvard Medical School (Boston, Massachusetts): Rohit Bakshi, Renxin Chu, Gloria Kim, Shahamat Tauhid, Subhash Tummala, Fawad Yousuf; Cedars-Sinai Medical Center (Los Angeles, California): Nancy L. Sicotte, Henry M. Jackson; Foundation for the Advancement of Military Medicine (Bethesda, Maryland): Dzung Pham, Snehashis Roy; National Institutes of Health (Bethesda, Maryland): Frances Andrada, Irene C.M. Cortese, Jennifer Dwyer, Rosalind Hayden, Haneefa Muhammad, Govind Nair, Joan Ohayon, Daniel S. Reich, Pascal Sati, Chevaz Thomas; Johns Hopkins (Baltimore, Maryland): Peter A. Calabresi, Sandra Cassard, Jiwon Oh; Oregon Health & Science University (Portland, Oregon): William Rooney, Daniel Schwartz, Ian Tagge; University of California (San Francisco, California): Roland G. Henry, Nico Papinutto, William Stern, Alyssa Zhu; University of Pennsylvania (Philadelphia, Pennsylvania): Christos Davatzikos, Jimit Doshi, Guray Erus, Kristin Linn, Russell Shinohara; University of Toronto (Toronto, Ontario, Canada): Jiwon Oh; Yale University (New Haven, Connecticut): R. Todd Constable, Daniel Pelletier.

Disclosures: Jordan D. Dworkin—RELATED: Grant: National Institutes of Health.* Kristin A. Linn—UNRELATED: Payment for Development of Educational Presentations: Institute of Electrical and Electronics Engineers. Ipek Oguz—RELATED: Grant: National Institutes of Health.* Rohit Bakshi—UNRELATED: Consultancy: EMD Serono, Genentech, Sanofi Genzyme, Novartis; Grants/Grants Pending: Biogen, EMD Serono, Novartis, Sanofi Genzyme.* Peter A. Calabresi—UNRELATED: Consultancy: Biogen, Disarm Therapeutics; Grants/Grants Pending: Genzyme, Annexon Bio, MedImmune*; Royalties: Cambridge Press, Comments: book on optical coherence tomography. Roland G. Henry—UNRELATED: Consultancy: Roche, MedDay; Grants/Grants Pending: Roche, Genentech, MedDay*; Payment for Lectures Including Service on Speakers Bureaus: Sanofi-Genzyme. Daniel Pelletier—UNRELATED: Consultancy: Genzyme Sanofi, Biogen, Novartis, Actelion, Genentech. William Rooney—UNRELATED: Employment: Oregon Health & Science University; Grants/Grants Pending: National Institutes of Health*; Payment for Lectures Including Service on Speakers Bureaus: Teva Pharmaceutical Industries; Patents (Planned, Pending or Issued): Pending: US 20150141804 A1, US 20150087967 A1; Issued: US 7286867 B2, US 9619875 B2.* Nancy L. Sicotte—RELATED: Grant: Race to Erase MS.* Russell T. Shinohara—RELATED: Grant: National Institutes of Health*; UNRELATED: Board Membership: Genentech, Comments: Scientific Advisory Board member; Expert testimony: Roche, Comments: legal consulting; Payment for Lectures Including Service on Speakers Bureaus: National Institutes of Health, National Multiple Sclerosis Society.* *Money paid to the institution.

REFERENCES

1. Radü EW, Sahraian MA, eds. *MRI Atlas of MS Lesions*. Berlin: Springer-Verlag; 2008

2. Barkhof F. **The clinico-radiological paradox in multiple sclerosis revisited.** *Curr Opin Neurol* 2002;15:239–45 CrossRef Medline
3. Popescu V, Agosta F, Hulst HE, et al; MAGNIMS Study Group. **Brain atrophy and lesion load predict long term disability in multiple sclerosis.** *J Neurol Neurosurg Psychiatry* 2013;84:1082–91 CrossRef Medline
4. Calabresi PA, Radue EW, Goodin D, et al. **Safety and efficacy of fingolimod in patients with relapsing-remitting multiple sclerosis (FREEDOMS II): a double-blind, randomised, placebo-controlled, phase 3 trial.** *Lancet Neurol* 2014;13:545–56 CrossRef Medline
5. Thompson AJ, Kermode AG, MacManus DG, et al. **Patterns of disease activity in multiple sclerosis: clinical and magnetic resonance imaging study.** *BMJ*. 1990;300:631–34 CrossRef Medline
6. Brex PA, Ciccarelli O, O'Riordan JL, et al. **A longitudinal study of abnormalities on MRI and disability from multiple sclerosis.** *N Engl J Med* 2002;346:158–64 CrossRef Medline
7. Khoury SJ, Guttmann CR, Orav EJ, et al. **Longitudinal MRI in multiple sclerosis: correlation between disability and lesion burden.** *Neurology* 1994;44:2120–24 CrossRef Medline
8. Rudick RA, Lee JC, Simon J, et al. **Significance of T2 lesions in multiple sclerosis: a 13-year longitudinal study.** *Ann Neurol* 2006;60:236–42 CrossRef Medline
9. Zivadinov R, Zorzon M, De Masi R, et al. **Effect of intravenous methylprednisolone on the number, size and confluence of plaques in relapsing-remitting multiple sclerosis.** *J Neurol Sci* 2008;267:28–35 CrossRef Medline
10. Harris JO, Frank JA, Patronas N, et al. **Serial gadolinium-enhanced magnetic resonance imaging scans in patients with early, relapsing-remitting multiple sclerosis: implications for clinical trials and natural history.** *Ann Neurol* 1991;29:548–55 CrossRef Medline
11. Guttmann CR, Rousset M, Roch JA, et al. **Multiple sclerosis lesion formation and early evolution revisited: a weekly high-resolution magnetic resonance imaging study.** *Mult Scler J* 2016;22:761–69 CrossRef Medline
12. Gaitán MI, Shea CD, Evangelou IE, et al. **Evolution of the blood-brain barrier in newly forming multiple sclerosis lesions.** *Ann Neurol* 2011;70:22–29 CrossRef Medline
13. Sweeney EM, Shinohara RT, Dewey BE, et al. **Relating multi-sequence longitudinal intensity profiles and clinical covariates in incident multiple sclerosis lesions.** *Neuroimage Clin* 2016;10:1–17 CrossRef Medline
14. Fonov V, Evans AC, Botteron K, et al; Brain Development Cooperative Group. **Unbiased average age-appropriate atlases for pediatric studies.** *Neuroimage* 2011;54:313–27 CrossRef Medline
15. Avants B, Tustison N, Song G, et al. *ANTS: Advanced Open-Source Normalization Tools for Neuroanatomy*. Philadelphia: Penn Image Computing and Science Laboratory; 2009
16. Carass A, Wheeler MB, Cuzzocreo J, et al. **A joint registration and segmentation approach to skull stripping.** In: *Proceedings of the 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Arlington, Virginia. April 12–15, 2007:656–59. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4193371. Accessed April 1, 2016
17. Shinohara RT, Sweeney EM, Goldsmith J, et al; Australian Imaging Biomarkers Lifestyle Flagship Study of Ageing, Alzheimer's Disease Neuroimaging Initiative. **Statistical normalization techniques for magnetic resonance imaging.** *Neuroimage Clin* 2014;6:9–19 CrossRef Medline
18. Sweeney EM, Shinohara RT, Shiee N, et al. **OASIS is Automated Statistical Inference for Segmentation, with applications to multiple sclerosis lesion segmentation in MRI.** *Neuroimage Clin* 2013;2:402–13 CrossRef Medline
19. Shinohara RT, Oh J, Nair G, et al; NAIMS Cooperative. **Volumetric analysis from a harmonized multisite brain MRI study of a single subject with multiple sclerosis.** *AJNR Am J Neuroradiol* 2017;38:1501–09 CrossRef Medline
20. Jenkinson M, Pechaud M, Smith S. **BET2: MR-based estimation of brain, skull and scalp surfaces.** Oxford: Oxford Centre for

- Functional Magnetic Resonance Imaging of the Brain (fMRIB); 2005. <http://mickaelpechaud.free.fr/these/HBM05.pdf>. Accessed June 13, 2016
21. Valcarcel AN, Linn KA, Vandekar SN, et al. MIMoSA: a method for inter-modal segmentation analysis. June 15, 2017. bioRxiv 150284 CrossRef
 22. Sweeney EM, Shinohara RT, Shea CD, et al. **Automatic lesion incidence estimation and detection in multiple sclerosis using multi-sequence longitudinal MRI.** *AJNR Am J Neuroradiol* 2013;34:68–73 CrossRef Medline
 23. Papinutto N, Bakshi R, Bischof A, et al; North American Imaging in Multiple Sclerosis Cooperative (NAIMS). **Gradient nonlinearity effects on upper cervical spinal cord area measurement from 3D T1-weighted brain MRI acquisitions.** *Magn Reson Med* 2018;79:1595–601 CrossRef Medline
 24. Oh J, Bakshi R, Calabresi PA; NAIMS Cooperative Steering Committee. **The NAIMS cooperative pilot project: Design, implementation and future directions.** *Mult Scler* 2017 Oct 1. [Epub ahead of print] CrossRef Medline
 25. Meier DS, Guttman CR. **Time-series analysis of MRI intensity patterns in multiple sclerosis.** *Neuroimage* 2003;20:1193–209 CrossRef Medline
 26. Meier DS, Guttman CR. **MRI time series modeling of MS lesion development.** *Neuroimage* 2006;32:531–37 CrossRef Medline