2014

Potholes and molehills: bias in the diagnostic performance of diffusion-tensor imaging in concussion

R. Watts

A. Thomas

C. G. Filippi
Zucker School of Medicine at Hofstra/Northwell

J. P. Nickerson

K. Freeman

Potholes and Molehills: Bias in the Diagnostic Performance of Diffusion-Tensor Imaging in Concussion[1]

Richard Watts, PhD
Alex Thomas, BA
Christopher G. Filippi, MD[2]
Joshua P. Nickerson, MD
Kalev Freeman, MD, PhD

Purpose: To investigate the extent of bias in a clinical study involving "pothole analysis" of diffusion-tensor imaging (DTI) data used to quantify white matter lesion load in diseases with a heterogeneous spatial distribution of pathologic findings, such as mild traumatic brain injury (TBI), and create a mathematical model of the bias.

Materials and Methods: Use of the same reference population to define normal findings and make comparisons with a patient group introduces bias, which potentially inflates reported diagnostic performance. In this institutional review board–approved prospective observational cohort study, DTI data were obtained in 20 patients admitted to the emergency department with mild TBI and in 16 control subjects. Potholes and molehills were defined as clusters of voxels with fractional anisotropy values more than 2 standard deviations below and above the mean of the corresponding voxels in the reference population, respectively. The number and volume of potholes and molehills in the two groups were compared by using a Mann-Whitney *U* test.

Results: Standard analysis showed significantly more potholes in mild TBI than in the control group (102.5 ± 34.3 vs 50.6 ± 28.9, $P < .001$). Repeat analysis by using leave-one-out cross-validation decreased the apparent difference in potholes between groups (mild TBI group, 102.5 ± 34.3; control group, 93.4 ± 27.2; $P = .369$). It was demonstrated that even with 100 subjects, this bias can decrease the voxelwise false-positive rate by more than 30% in the control group.

Conclusion: The pothole approach to neuroimaging data may introduce bias, which can be minimized by independent training and test groups or cross-validation methods. This bias is sufficient to call into question the previously reported diagnostic performance of DTI for mild TBI.

Online supplemental material is available for this article.

[1] From the Departments of Radiology (R.W., C.G.F., J.P.N.), Surgery (A.T., K.V.), and Neurology (C.G.F.), University of Vermont, Given Medical Building E301, 89 Beaumont Ave, Burlington, VT 05405. Received August 17, 2013; revision requested September 23; revision received November 18; accepted December 3; final version accepted January 9, 2014. Supported by the Totman Medical Research Trust and grants from the U.S. Department of Energy (SC 0001753) and Department of Defense (W911 NF-10-1-0376). **Address correspondence to** K.F. (e-mail: *kalev.freeman@uvm.edu*).

[2] **Current address:** Department of Radiology, Columbia University Medical Center, New York, NY.

Diagnosing mild traumatic brain injury (TBI) after concussion by using brain imaging is fundamentally a difficult problem; it requires that the effect of a single noncatastrophic event can be recognized as having distinct characteristics, as opposed to those caused by natural variation among the population amid a lifetime background of other minor insults. Further confounding this problem is the inherent heterogeneity of mild TBI, as the spatial distribution and magnitude of any effect are likely to vary markedly from one individual to another (1).

In severe TBI, diffuse axonal injury, along with other types of neurotrauma, are commonly diagnosed by using magnetic resonance (MR) imaging. The effects of mild TBI are too subtle for such qualitative analysis, but diffusion-tensor imaging (DTI) has shown promise, with recent publications suggesting that quantitative analysis of white matter fractional anisotropy (FA) could serve as a diagnostic modality for mild TBI (2–10). Given the heterogeneity of the trauma to the brain in mild TBI and the variable spatial distribution and magnitude of damage, it seems naive to expect that an average value of FA within a specific, large region of white matter would have diagnostic value. The range of damage would make the distribution of values broad within the patients with mild TBI, while natural variability will add another confounding factor. If the effects of the mild TBI are focal, then the detection power will be limited by averaging values with those of unaffected tissue. Several prior studies (11–15), though not all (16), have shown group differences between patients with mild TBI and control subjects in region of interest analysis, but the diagnostic performance characteristics of DTI for mild TBI are largely unknown.

An alternative approach is that of "pothole" analysis. White et al introduced the concept of analyzing white matter potholes in a study of early-onset schizophrenia (17), and this technique has been applied recently to white matter analysis in mild TBI (2,4,7–10). In this technique, the FA value at each voxel is transformed into a z statistic based on the mean and the standard deviation of FA in a reference population. White matter potholes are defined as clusters of voxels in which the FA z statistic is below some threshold. We similarly define molehills as clusters of voxels with increased FA. Additional constraints, such as defining a minimum cluster size, may be applied. The summary statistic may be the total number of such clusters or the total volume of clusters.

We hypothesized that the total number of potholes or molehills would be higher in patients with mild TBI than in a control group. We replicated previously published methods for pothole analysis (2,7–9), in which images from control subjects were used both for generation of a reference standard and for comparison with patients with mild TBI, and we compared this by using leave-one-out cross-validation (18). We sought to investigate the extent of this bias in a prospective observational cohort study, create a mathematical model of the bias, and apply this model to prior published studies of mild TBI.

Advances in Knowledge

- In diffusion-tensor imaging (DTI) analysis, "potholes" are defined as clusters of voxels with reduced fractional anisotropy (FA) values compared with the corresponding voxels in a reference population, and pothole analysis may provide a useful biomarker for mild traumatic brain injury (TBI); however, the use of nonindependent data for both reference and comparison to an independent mild TBI group may introduce bias.
- We have demonstrated in an experimental study that nonindependence of the reference and comparison populations can produce highly significant differences in the number of FA "potholes" ($P < .001$) between mild TBI and control groups, which fail to reach significance ($P > .05$) by using unbiased cross-validation.

Implications for Patient Care

- The diagnostic utility of DTI "pothole" analysis as reported in the literature may be overly optimistic, owing to bias in the analysis.
- Caution should be exercised in transitioning these techniques to clinical practice.

Materials and Methods

Study Design and Population

We performed a prospective observational cohort study that was approved by the University of Vermont institutional review board, with all subjects having given written consent. We enrolled men and women aged 18–57 years with mild TBI who were admitted to the emergency department at the University of Vermont/Fletcher-Allen Health Care center within 72 hours of injury. Specific inclusion criteria for mild TBI were acute head injury, Glasgow Coma Scale score of 13–15, and two or more concussive symptoms (loss of consciousness, blurred vision, confusion, dizziness, memory problems, or poor balance). Head computed

Published online before print
10.1148/radiol.14131856 **Content codes:** MR NR

Radiology 2014; 272:217–223

Abbreviations:
DTI = diffusion-tensor imaging
FA = fractional anisotropy
TBI = traumatic brain injury

Author contributions:
Guarantors of integrity of entire study, R.W., K.F.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, all authors; clinical studies, R.W., A.T., C.G.F., K.F.; experimental studies, A.T., C.G.F., K.F.; statistical analysis, R.W. A.T., K.F.; and manuscript editing, all authors

Funding:
This research was supported by the National Institutes of Health (grants K08 GM 098795-01, GM 098795, and HL 120877).

Conflicts of interest are listed at the end of this article.

tomographic (CT) scans were obtained the discretion of the provider, and a clinical report was generated by the attending neuroradiologist. Exclusion criteria were *(a)* moderate to severe TBI, requiring acute neurosurgical intervention or hospitalization; *(b)* concomitant injuries (defined as an injury severity score for any other organ system > 2); *(c)* history of disabling TBI (defined as a prior head injury with persistent postconcussive symptoms; we did not exclude patients with prior mild TBI who reported full recovery); *(d)* preexisting neurologic disorder or psychiatric condition that required medical treatment within the past year; and *(e)* contraindications to MR imaging. Patients were also excluded if they were enrolled initially but MR images were not obtained within 72 hours of injury.

The control group consisted of healthy volunteers without acute injury who responded to flier advertisements or patients with extremity injuries that were admitted to the emergency department within 72 hours of injury. Extremity injuries were defined as an isolated injury to either the arms or legs and no head trauma or TBI symptoms. Research staff completed the initial assessment by reviewing the hospital chart, having a discussion with the subject's provider, and conducting a structured interview of the subject while he or she was in the emergency department.

Outcome Measures and Assessment

Initial brain MR examinations were completed within 72 hours of injury, and follow-up images were completed 7–10 days after injury. Images were acquired with a Philips Achieva TX 3.0-T unit (Philips Healthcare, Best, the Netherlands) by using an eight-channel brain coil. T1-weighted images were acquired by using a three-dimensional inversion-recovery spoiled gradient-echo technique: repetition time (msec)/echo time (msec)/inversion time (msec), 8.1/3.7/1008; flip angle, 8°; and sensitivity encoding factor, 1.5. A sagittal acquisition matrix of 240 × 240 × 160 provided whole-brain coverage, with an isotropic 1-mm spatial resolution and an imaging time of less than 8 minutes. Fluid-attenuated T2-weighted images were acquired by using three-dimensional fluid-attenuated inversion-recovery technique, with a sagittal field of view of 250 × 250 × 180 mm, an acquisition matrix of 224 × 224 × 160 to give isotropic 1.1-mm resolution, 4800/279/1650, and an imaging time of less than 5 minutes. To rule out hemorrhage, susceptibility-weighted images were acquired by using a three-dimensional T2*-weighted gradient-echo technique (principles of echo shifting with a train of observations) with 15/21 (echo time shifted). An axial acquisition matrix of 220 × 180 × 100 was used with 1-mm isotropic resolution and an imaging time of less than 5 minutes.

Diffusion-weighted images were acquired by using a single-shot spin-echo echo-planar imaging acquisition with a b value of 1000 sec/mm^2 and 46 uniformly distributed noncollinear directions. An additional six images were acquired with no diffusion weighting (b = 0 sec/mm^2). The acquisition matrix was 120 × 120, with a field of view of 240 × 240 mm by using a sensitivity encoding factor of two. Fifty-nine contiguous 2-mm-thick sections were acquired and were aligned to the anterior commissure and posterior commissure axis, with 10000/68 and an imaging time of 9 minutes. Details of image processing are provided in Appendix E1 (online).

All MR images were reviewed by a board-certified neuroradiologist (C.G.F., with 16 years of experience, or J.P.N., with 3 years of experience, both with Certificates of Added Qualification for neuroradiology) to identify lesions both relating to and unrelated to trauma.

Statistical Analysis

Standard pothole analysis.—We used previously published methods for quantifying the numbers of potholes in mild TBI (2,7–9). We defined potholes as clusters of voxels larger than 30 mm^3 in which the z statistic was below −2; we similarly defined molehills as having a z statistic higher than +2. For this analysis, both the patients with mild TBI (independent group) and control subjects (dependent group) were analyzed on the basis of the voxelwise mean and standard deviation values derived from the control subjects at the first time point. Both the number and total volume of the potholes and molehills identified were calculated for each subject and used as summary statistics for group comparison.

Leave-one-out cross-validation analysis.—To distinguish the training data set used to estimate the mean and standard deviation of the healthy population from the testing data set of control subjects used for comparison with patients with mild TBI, we used leave-one-out cross-validation (18). Each round of cross-validation serves to partition the data by separating results for the control subject to be analyzed from the mean and standard deviation calculations used to create the z statistic for the control population. Thus, each control subject was analyzed by using a reference group consisting of the other control subjects.

We then generalized the leave-one-out analysis (Appendix E2 [online]) to calculate the effective z statistic and the ratio of false-positive findings in the control group as compared with the patient group for any combination of subject number and threshold z value. A simple spreadsheet is also provided (Fig E1 [online]) to enable the reader to investigate this effect numerically.

Numerical calculation of methodological bias.—To clarify the discrepancy in results between standard pothole analysis and leave-one-out cross-validation approaches, we calculated the degree of bias introduced by nonindependent reference and control groups with different sample sizes analytically.

All statistical calculations were performed with SPSS software (SPSS Statistics for Windows, version 20.0; IBM, Armonk, NY), and values were given as means ± standard deviations, unless noted otherwise. Group comparisons were performed with Mann-Whitney U tests for nonparametric data or two-sample t tests, as noted.

All quantitative analyses were performed by R.W. and A.T. (under the supervision of R.W.). R.W. is an MR

imaging physicist with 15 years of experience in MR imaging data acquisition and analysis.

Results

Enrollment of Patients with Mild TBI and Control Subjects

We initially enrolled 28 patients with mild TBI and 20 control subjects, but eight patients with mild TBI and four control subjects were excluded because we were unable to obtain adequate images within 72 hours of injury (seven had the incorrect DTI sequence performed; three were unable to make it to MR imaging in time; and two had excessive movement in the MR imager). We ultimately included 20 patients with mild TBI and 16 control subjects in our analysis. Control subjects included seven volunteers without acute injuries or history of brain trauma and nine patients in the emergency department with extremity injuries and absence of head trauma. Control subjects were not age or sex matched but were rather chosen by means of random selection. Subject demographics are shown in Table 1. There were no significant differences between the control group and the mild TBI population with regard to age, sex, handedness, or education. Among the 20 patients with mild TBI, the treating physician performed CT in 12; one patient had a subtle, small area of intraparenchymal hemorrhage. This subject was not excluded. Findings in the remaining 11 CT examinations were interpreted as being normal.

Total Numbers of Potholes and Molehills

No additional focal lesions were identified at qualitative radiologic review of MR images. Standard quantitative pothole analysis results are shown in Figure, *A*, and Table 2. There was a large, significant difference in the number of FA potholes between the mild TBI group and the control population (102.5 ± 34.3 vs 50.6 ± 28.9, respectively; $P < .001$). We then performed a leave-one-out cross-validation approach to establish an independent reference group with which to compare control subjects and patients with mild TBI (Fig, *B*, Table 2). When the leave-one-out method is used in the same subjects, the difference in numbers of potholes between the control subjects and patients with mild TBI loses significance (102.5 ± 34.3 vs 93.4 ± 27.2; $P = .369$). Similarly, highly significant differences were also seen between patients with mild TBI and control subjects when looking at total numbers of molehills or total volumes of either potholes or molehills at either of the two time points studied (Fig E2

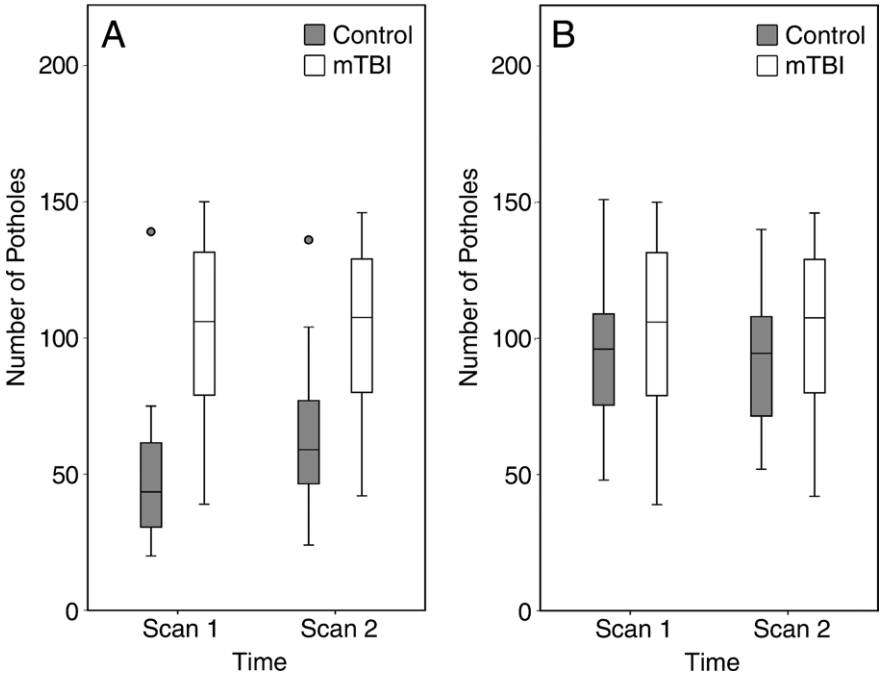
Table 1

Demographic Information for Patients with Mild TBI in the Emergency Department and Control Subjects

Parameter	Mild TBI Group	Control Group
No. of subjects	20	16
No. of men	11 (55)	7 (44)
No. of subjects with right-handedness	17 (85)	13 (81)
Mean age (y)	30.6 ± 12	28.1 ± 9.4
Mean length of education (y)	14.7 ± 2.0	15.7 ± 2.4
Mean time from injury to first MR imaging examination (d)	1.9 ± 0.9	2.4 ± 0.5*
Mean time from injury to second MR imaging examination (d)	8.6 ± 1.3	9.3 ± 1.6*
Mean time between MR imaging examinations (d)	6.7 ± 1.1	6.9 ± 1.7

Note.—Data are either numbers of patients, with percentages in parentheses, or means ± standard deviations.

* Values are given for the nine control subjects with trauma only.



A, Box and whisker plot shows comparison of patients with mild TBI *(mTBI)* vs control subjects less than 72 hours after injury and 1 week after injury by using biased pothole analysis. The reference population consists of the control group at the first time point. *B*, Box and whisker plot demonstrates corresponding analysis by using leave-one-out cross-validation, which eliminates the bias due to nonindependence of the reference population, resulting in no significant difference between the groups.

Table 2

Results of Mann-Whitney *U* Tests to Compare Standard (Biased) and Unbiased Leave-One-Out Analysis of an Independently Collected Prospective Study of Patients with Mild TBI Imaged Up to 72 Hours after Injury and 1 Week after Injury

		Standard (Biased) Analysis		Leave-One-Out Cross-Validation	
Imaging Time and Parameter	Mild TBI Group	Control Group	*P* Value	Control Group	*P* Value
Imaging at <72 hours after injury					
Potholes					
No.	102.5	50.6	<.001	93.4	.369
Volume	12065	4842	<.001	11705	.604
Molehills					
No.	90.8	46.5	<.001	86.5	.604
Volume	12550	4279	<.001	11583	.336
Imaging at 1 week after injury					
Potholes					
No.	102.5	62.9	.001	93.0	.404
Volume	11856	6643	.001	12036	.604
Molehills					
No.	87.0	65.3	.039	91.2	.694
Volume	12716	6806	.005	12121	.626

Note.—The reference group consisted of the control subjects imaged at the first time point for both analyses. Identical results for imaging up to 72 hours after injury were obtained by applying a reduced *z* threshold of 1.724 to the biased *z* statistics. Repeating the analysis by using a two-sample *t* test also yielded significant differences in all the biased analyses ($P < .05$) and no significant differences in the leave-one-out analysis.

[online]). Again, these differences disappeared when using the leave-one-out method (Table 2).

Example of Methodological Bias Introduced by Pothole Analysis

Consider a study in which 15 reference subjects are used to estimate the mean and standard deviation of the healthy population. The *z* score obtained at a particular voxel for a 16th subject will vary, depending on whether this subject is added to the reference population or not. We accept a *z* score of −2 or less as constituting a pothole. The mean and unbiased (σ_{n-1}) estimates of FA from the 15 control subjects are calculated to be 0.500 ± 0.050. In our new subject, an FA value of 0.390 is observed. Our *z* statistic is then calculated as:

$$z = \frac{X - \mu}{\sigma} = \frac{0.390 - 0.500}{0.050} = -2.20,$$

where *X* represents the FA value for this subject. This meets our criteria to be classified as a pothole, with a corresponding *P* value of .014.

If we now include this subject in our estimate of the mean and standard deviation of the reference population, then the biased mean and standard deviation become 0.493 ± 0.056. Repeating the calculation results in a *z* statistic of −1.85 and a *P* value of .032. This voxel would no longer be classified as a pothole. In this case, to achieve an apparent significance level, *z* less than −2 requires an FA value of 0.378 or less, with a true (unbiased) *z* statistic of −2.44 or less. Because of the nonlinear relationship between *z* statistic and *P* value, we are likely to see three times fewer false-positive findings in the dependent group (control subjects) compared with the independent group (patients with mild TBI) in the absence of any true effect. Cluster analysis introduces a further nonlinear relationship, which would amplify this effect.

We provide analytical (Appendix E2 [online]) and numerical (Fig E1 [online]) calculations that demonstrate the bias for different sample sizes. Some representative examples are shown in Table 3. It can be seen that including the subject to be analyzed in the reference population results in a dramatic reduction in the number of significant voxels. Even with 100 subjects and a *z* threshold of 3, the number of significant voxels in the control group decreases by more than 30%, which is more than sufficient to result in a group difference between patients with mTBI and healthy control subjects.

Discussion

Our major finding is the identification of an important but largely unrecognized source of bias in many articles in which a pothole and molehill approach is used, owing to nonindependence of the reference population used to define normal findings and the control population used for comparison with the patient group. We demonstrated the practical difference in bias by using leave-one-out cross-validation analysis with independently collected, prospective MR imaging data in patients with mild TBI. We showed that for commonly used sample sizes, the bias introduced by pothole analysis is large. Use of the same data for both the reference and control groups effectively applies a higher-threshold *z* statistic to the dependent control group compared with the independent group. This can be avoided by the use of a separate, independent reference group or by suitable adjustment of the threshold *z* statistic.

Our findings are important because they suggest that the diagnostic utility of DTI "pothole" analysis as reported in the literature may be overly optimistic, owing to bias in the analysis that effectively decreases the number of both molehills and potholes in the control group; this bias may be minimized in future studies by using the corrected *z* statistic threshold or the leave-one-out method. Our experimental study also shows that the bias is largely maintained at the second time point, despite the independence of the data acquisition in this case (the reference population was the source of the control data

Table 3

Bias Introduced by Including the Subject to Be Analyzed in the Reference Sample to Define the Mean and Standard Deviation Used to Calculate the *z* Statistic

Study	No. of Independent Control Subjects	Z_{Ind}	$Z_{Eff\text{-}Dep}$	$Z_{Corr\text{-}Dep}$	$P(Z_{Ind})/P(Z_{Eff\text{-}Dep})$
Mayer et al (9)	14	2.00	2.48	1.71	.29
Present study	15	2.00	2.44	1.72	.32
Jorge et al (2)	20	3.00	4.23	2.43	.0087
Ling et al (8)	49	2.00	2.11	1.90	.77
Present study (model)	49	3.00	3.36	2.73	.29
Present study (model)	100	3.00	3.16	2.86	.58

Note.—Z_{Ind} is the *z* statistic threshold applied to the independent mild TBI data. $Z_{Eff\text{-}Dep}$ is the corresponding effective *z* statistic threshold erroneously applied to the dependent control data set (the true *z* value required to achieve significance in the biased analysis). $Z_{Corr\text{-}Dep}$ is the corrected *z* statistic threshold that should be applied to the control data to produce the same result as the leave-one-out method with a threshold of Z_{Ind}. $P(Z_{Ind})/P(Z_{Eff\text{-}Dep})$ is the ratio of the corresponding *P* values (false-positive rates).

acquired at the first time point). This implies that the intersubject variability is much greater than the measurement error. Using the metric of FA in DTI analysis may therefore be subject to a ceiling in sensitivity and specificity, owing to intrinsic subject variability in cross-sectional studies. Alternative acquisition and analysis strategies (eg, those presented in references 19–22) may produce metrics that are more specific to mild TBI, although in general, these strategies require substantially increased imaging times that may be prohibitive in patients with acute TBI.

We identified four recently published studies in which significant differences were reported in potholes between control subjects and patients with mild TBI by using nonindependent control and reference groups (2,7–9). First, Jorge et al applied the pothole method to a study of 72 veterans with mild TBI related to blast exposure, 21 veterans without blast exposure, and 14 civilian patients with mild TBI (2). The 21 veterans without blast exposure were chosen as the reference population. A second study by Davenport et al (7) was conducted in a similar population, but the veterans with and those without blast-related mild TBI were segregated according to history of civilian mild TBI. Fourteen veterans with no history of mild TBI were used as the reference population. In that study, the nonblast group included the reference population. Third, Mayer et al (9) studied pediatric mild TBI and found that metrics of increased anisotropy were able to allow objective classification of pediatric mild TBI cases and healthy control subjects with 90% accuracy on the basis of a study of 15 pediatric patients with semiacute mild TBI, aged 10–17 years, and 15 matched control subjects, which were also used as the reference population. Finally, Ling et al described a pooled study of 50 adult patients with mild TBI and 50 matched control subjects (8), including pothole analysis, with the control subjects used as the reference population. Comparison of the latter two studies provides an interesting observation. Both studies are from the same group, and similar methods were used. Despite the much greater statistical power afforded in the adult study, the group difference in the number of clusters was only mildly significant ($P = .012$), while in the pediatric study, it was highly significant ($P < .00001$). While we cannot exclude that these populations respond very differently to mild TBI, it seems likely that this disparity is due to the reduced bias introduced when a larger reference population is used.

There are only two mild TBI studies that explicitly address the bias that results from pothole analysis of MR imaging data. Lipton et al (10) published a study in which they claimed that DTI allows for the robust detection of traumatic axonal injury in individual patients with mild TBI, in which there was a comparison between 34 patients with mild TBI and 30 control subjects by using a pothole-based method, EZ-MAP (10). The investigators found both potholes and molehills in the mild TBI group, but numbers of potholes were not reported for the control group. A subsequent study by Kim et al (4) explicitly used an independent control group and determined performance characteristics of both EZ-MAP (sensitivity, 71%; specificity, 71%) and conventional pothole methods (sensitivity, 65%; specificity, 76%). Of interest, none of the investigators in these studies reported the analyses of both molehills and potholes, with the notable exception of the study by Kim et al (4), which we speculate may be because the data appeared inconsistent with a simple model of FA increases or decreases.

Limitations of our experimental study included the use of a relatively small number of patients, which increases the bias identified here. We did not exclude patients with prior mild TBI who recovered fully. The use of a control group, excluding any subjects with a remote history of head injury or concussion, might have yielded different results. Other investigators have used different selection criteria and definitions of mild TBI, which may have included more severe trauma. However, our study population may be more generalizable to emergency medicine clinical practice.

In summary, the pothole and molehill approach to the analysis of DTI data is a potentially useful method that can be used to avoid many of the problems of traditional region of interest–based methods, which improves the detection effectiveness for any disease process with a heterogeneous spatial distribution of pathologic findings. However, care must be taken to avoid bias, and an explicit statement about the independence of the training and test groups should be required. The use of nonindependent reference and control groups in "pothole" analysis has led to a substantial overestimation of the

diagnostic utility of DTI for mild TBI in much of the literature. More studies are needed to determine measures such as sensitivity, specificity, positive predictive value, and negative predictive value. These studies should ideally be blinded multicenter trials to unambiguously demonstrate the independence of the testing data set and establish whether normal values can be generalized across sites.

Acknowledgments: The authors thank Trevor Andrews, Steve Braff, Warren Lockette, Magdalena Naylor, and Marina Shpaner for useful discussions. We are also grateful to emergency department staff and research associates for recruiting the subjects and to MR imaging technicians Jay Gonyea and Scott Hipko for their dedication and flexibility in imaging the subjects, often at short notice. Finally, we thank the patients and volunteers who made this study possible.

Disclosures of Conflicts of Interest: R.W. No relevant conflicts of interest to disclose. **A.T.** No relevant conflicts of interest to disclose. **C.G.F.** No relevant conflicts of interest to disclose. **J.P.N.** No relevant conflicts of interest to disclose. **K.F.** No relevant conflicts of interest to disclose.

References

1. Rosenbaum SB, Lipton ML. Embracing chaos: the scope and importance of clinical and pathological heterogeneity in mTBI. Brain Imaging Behav 2012;6(2):255–282.
2. Jorge RE, Acion L, White T, et al. White matter abnormalities in veterans with mild traumatic brain injury. Am J Psychiatry 2012; 169(12):1284–1291.
3. Kasahara K, Hashimoto K, Abo M, Senoo A. Voxel- and atlas-based analysis of diffusion tensor imaging may reveal focal axonal injuries in mild traumatic brain injury—comparison with diffuse axonal injury. Magn Reson Imaging 2012;30(4):496–505.
4. Kim N, Branch CA, Kim M, Lipton ML. Whole brain approaches for identification of microstructural abnormalities in individual patients: comparison of techniques applied to mild traumatic brain injury. PLoS ONE 2013; 8(3):e59382.
5. Lipton ML, Gulko E, Zimmerman ME, et al. Diffusion-tensor imaging implicates prefrontal axonal injury in executive function impairment following very mild traumatic brain injury. Radiology 2009;252(3):816–824.
6. Matsushita M, Hosoda K, Naitoh Y, Yamashita H, Kohmura E. Utility of diffusion tensor imaging in the acute stage of mild to moderate traumatic brain injury for detecting white matter lesions and predicting long-term cognitive function in adults. J Neurosurg 2011;115(1):130–139.
7. Davenport ND, Lim KO, Armstrong MT, Sponheim SR. Diffuse and spatially variable white matter disruptions are associated with blast-related mild traumatic brain injury. Neuroimage 2012;59(3):2017–2024.
8. Ling JM, Peña A, Yeo RA, et al. Biomarkers of increased diffusion anisotropy in semiacute mild traumatic brain injury: a longitudinal perspective. Brain 2012;135(Pt 4): 1281–1292.
9. Mayer AR, Ling JM, Yang Z, Pena A, Yeo RA, Klimaj S. Diffusion abnormalities in pediatric mild traumatic brain injury. J Neurosci 2012;32(50):17961–17969.
10. Lipton ML, Kim N, Park YK, et al. Robust detection of traumatic axonal injury in individual mild traumatic brain injury patients: intersubject variation, change over time and bidirectional changes in anisotropy. Brain Imaging Behav 2012;6(2):329–342.
11. Bazarian JJ, Zhong J, Blyth B, Zhu T, Kavcic V, Peterson D. Diffusion tensor imaging detects clinically important axonal damage after mild traumatic brain injury: a pilot study. J Neurotrauma 2007;24(9):1447–1459.
12. Inglese M, Makani S, Johnson G, et al. Diffuse axonal injury in mild traumatic brain injury: a diffusion tensor imaging study. J Neurosurg 2005;103(2):298–303.
13. Mac Donald CL, Johnson AM, Cooper D, et al. Detection of blast-related traumatic brain injury in U.S. military personnel. N Engl J Med 2011;364(22):2091–2100.
14. Mayer AR, Ling J, Mannell MV, et al. A prospective diffusion tensor imaging study in mild traumatic brain injury. Neurology 2010; 74(8):643–650.
15. Miles L, Grossman RI, Johnson G, Babb JS, Diller L, Inglese M. Short-term DTI predictors of cognitive dysfunction in mild traumatic brain injury. Brain Inj 2008;22(2): 115–122.
16. Lange RT, Iverson GL, Brubacher JR, Mädler B, Heran MK. Diffusion tensor imaging findings are not strongly associated with postconcussional disorder 2 months following mild traumatic brain injury. J Head Trauma Rehabil 2012;27(3):188–198.
17. White T, Schmidt M, Karatekin C. White matter 'potholes' in early-onset schizophrenia: a new approach to evaluate white matter microstructure using diffusion tensor imaging. Psychiatry Res 2009;174(2):110–115.
18. Picard RR, Cook RD. Cross-validation of regression-models. J Am Stat Assoc 1984;79 (387):575–583.
19. Raffelt D, Tournier JD, Rose S, et al. Apparent Fibre Density: a novel measure for the analysis of diffusion-weighted magnetic resonance images. Neuroimage 2012;59(4): 3976–3994.
20. Zhuo J, Xu S, Proctor JL, et al. Diffusion kurtosis as an in vivo imaging marker for reactive astrogliosis in traumatic brain injury. Neuroimage 2012;59(1):467–477.
21. Fernandez-Miranda JC, Pathak S, Engh J, et al. High-definition fiber tractography of the human brain: neuroanatomical validation and neurosurgical applications. Neurosurgery 2012;71(2):430–453.
22. Yeh FC, Wedeen VJ, Tseng WY. Generalized q-sampling imaging. IEEE Trans Med Imaging 2010;29(9):1626–1635.
23. Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2(8):e124.