Methods for Identifying Subject-Specific Abnormalities in Neuroimaging Data

Andrew R. Mayer,[1,2,3]* Edward J. Bedrick,[4] Josef M. Ling,[1] Trent Toulouse,[1] and Andrew Dodd[1]

[1]*The Mind Research Network/Lovelace Biomedical and Environmental Research Institute, Albuquerque, New Mexico*
[2]*Neurology Department, University of New Mexico School of Medicine, Albuquerque, New Mexico*
[3]*Department of Psychology, University of New Mexico, Albuquerque, New Mexico*
[4]*Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, Colorado*

Abstract: Algorithms that are capable of capturing subject-specific abnormalities (SSA) in neuroimaging data have long been an area of focus for diverse neuropsychiatric conditions such as multiple sclerosis, schizophrenia, and traumatic brain injury. Several algorithms have been proposed that define SSA in patients (i.e., comparison group) relative to image intensity levels derived from healthy controls (HC) (i.e., reference group) based on extreme values. However, the assumptions underlying these approaches have not always been fully validated, and may be dependent on the statistical distributions of the transformed data. The current study evaluated variations of two commonly used techniques ("pothole" method and standardization with an independent reference group) for identifying SSA using simulated data (derived from normal, *t* and chi-square distributions) and fractional anisotropy maps derived from 50 HC. Results indicated substantial group-wise bias in the estimation of extreme data points using the pothole method, with the degree of bias being inversely related to sample size. Statistical theory was utilized to develop a distribution-corrected *z*-score (DisCo-Z) threshold, with additional simulations demonstrating elimination of the bias and a more consistent estimation of extremes based on expected distributional properties. Data from previously published studies examining SSA in mild traumatic brain injury were then re-analyzed using the DisCo-Z method, with results confirming the evidence of group-wise bias. We conclude that the benefits of identifying SSA in neuropsychiatric research are substantial, but that proposed SSA approaches require careful implementation under the different distributional properties that characterize neuroimaging data. *Hum Brain Mapp 00:000–000, 2014.* © **2014 Wiley Periodicals, Inc.**

Key words: mild traumatic brain injury; diffusion tensor imaging; neuroimaging; subject-specific abnormalities

INTRODUCTION

Diffusion tensor imaging (DTI) studies are increasingly being used to characterize white matter abnormalities in neuropsychiatric populations including multiple sclerosis (Ge et al., 2005), schizophrenia (Davis et al., 2003; Scheel et al., 2012), mild traumatic brain injury (mTBI; Hulkower et al., 2013; Niogi and Mukherjee, 2010; Shenton et al., 2012), and substance abuse disorders (Monnig et al., 2013). The majority of published studies utilize traditional region

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Andrew Mayer, The Mind Research Network, Pete & Nancy Domenici Hall, 1101 Yale Blvd. NE, Albuquerque, NM 87106. E-mail: amayer@mrn.org

Received for publication 17 December 2013; Revised 20 May 2014; Accepted 27 May 2014.

DOI: 10.1002/hbm.22563

Published online 00 Month 2014 in Wiley Online Library (wileyonlinelibrary.com).

of interest (ROI) or voxel-wise analyses to directly compare patient groups to healthy controls (HC). This analytic approach is based on the implicit assumption that clinically heterogeneous patients have a homogenous pattern of image-based abnormalities (i.e., high degree of spatial overlap), which in turn results in statistically different sample means during voxel-wise or ROI comparisons. However, it is increasingly recognized that the pattern of white matter injury may vary both across individual patients in mTBI (Bouix et al., 2013; Kim et al., 2013; Ling et al., 2012; Lipton et al., 2012; Pasternak et al., 2014; White et al., 2009), as well as change as a function of disease progression. The latter point is most readily apparent for multiple sclerosis, in which white matter lesions occur and remit in different locations as a natural function of the disease course (Ge et al., 2005).

The need for developing automated algorithms that classify subject-specific abnormalities (SSA) on a voxel-wise basis has long been recognized for both clinical and research purposes (Poline and Mazoyer, 1993). The first step for identifying a SSA is to compare an individual subject's data to a known reference distribution, typically derived from a cohort of HC (Viviani et al., 2007). SSA are typically determined based on an extreme deviation from a known or estimated probability density function. However, a percentage of SSA is expected based on chance alone even in HC, and researchers typically wish to determine whether the frequency of abnormalities is statistically different in patient samples relative to controls. A spatially variable pattern of SSA (i.e., scattered lesions) would likely result in only small (i.e., nonsignificant) deviations (e.g., slight skew or kurtosis) in the distribution of patient and control data on a voxel-wise level. Therefore, two stages of data analyses are typically necessary for determining (1) whether the subject-specific data points are abnormal (i.e., extreme), and (2) whether the frequency of SSA (similar to a metric of lesion load) is statistically different across samples of HC and patients.

The "pothole" method (White et al., 2009) has proven to be popular in the neuroimaging community (Davenport et al., 2012; Ehrlich et al., 2013; Jorge et al., 2012; Ling et al., 2012; Mayer et al., 2012; White et al., 2013) for comparing the number of SSA between patients (hereafter referred to as the comparison group) and HC (hereafter referred to as the reference group) across several different neuroimaging modalities. In the pothole method, data from both the reference and comparison groups are z-transformed on a voxel-wise basis using statistical moments (mean and standard deviation) derived from the reference group. The z-transformed voxels are subsequently classified as being normal or abnormal based on study specific thresholds (e.g., ~95th percentile using $z < -2$ or $z > 2$), typically in conjunction with a spatial-volume threshold (e.g., 20 contiguous voxels) to reduce the likelihood of false positives, similar to other cluster-based approaches frequently used in the literature (Friston et al., 1996; Hayasaka et al., 2004).

Other methods for identifying SSA include z-transforming each member of the reference group on the basis of all other members from the reference group (leave-one-out sampling), while z-transforming the comparison group based on all members of the reference group (Bouix et al., 2013; Pasternak et al., 2014). In this approach, all data are therefore z-transformed using various independent samples but with identical z-thresholds. Similar cross-validation techniques have been used to determine significant departures from the reference distribution in individual patients with various neurological disorders based on *t* random fields (Viviani et al., 2007).

Bootstrapping is another popular technique for estimating extreme values (Viviani et al., 2007). Bootstrapping has been used in conjunction with the z-transform approach (Enhanced *Z*-score Microstructural Assessment of Pathology [EZ-MAP]) to reduce the impact of variance scaling for both between-subject (Kim et al., 2013) and within-subject (Lipton et al., 2012) comparisons following mTBI. Importantly, to date the EZ-MAP method has been implemented using an independent sample of controls as the reference group, which is statistically similar to the leave-one-out approach followed by an additional correction for variance scaling (Bouix et al., 2013; Pasternak et al., 2014). Other approaches for determining extreme values are the one versus many *t*-test (Lipton et al., 2008; Patel et al., 2007) and the use of binomial distributions (Mac Donald et al., 2011). Finally, machine learning algorithms have been used to identify traumatic axonal injury using microbleeds as priors (Hellyer et al., 2012).

The underlying assumptions of all approaches (e.g., the impact of sample size and distribution properties) have not always been thoroughly considered in the context of typical neuroimaging data (Viviani et al., 2007), which can have critical implications and potentially lead to bias (i.e., a systematic difference across repeated samples between a sample statistic and a population parameter). For example, the pothole approach inherently assumes that the resultant distributions of the z-transformed data are identical across both the reference and comparison groups. However, differences in variance exist for z-transformed data dependent on whether the data point is included in the computation of the reference mean. Consequently, the statistical distributions of the z-transforms are different for the reference and comparison groups even in the null setting where the population distributions are identical (i.e., when two samples of HC are derived from the overall population). In contrast to previously published methods such as cross-validation and the EZ-MAP (Bouix et al., 2013; Kim et al., 2013), an alternative approach is to use statistical theory to derive distribution-corrected *z*-score (DisCo-Z) adjustments for the thresholds of the reference and comparison groups so that the probability of extreme values is identical.

To this end, we provide a theoretical consideration of the probability distributions of z-transformed data from the reference (HC) and comparison (patient group) samples. We assume that each observation in a given voxel/ROI is standardized by the mean $\bar{X}$ and standard

deviation s from the reference group and that the reference and comparison populations have identical multivariate normal (i.e., Gaussian) distributions of responses across voxels. For a given voxel, let $Z_R=(X_R-\bar{X})/s$ and $Z_C=(X_C-\bar{X})/s$ be the z-score transforms for random responses X_R and X_C from the reference and comparison populations. The reference group mean $\bar{X}$ and standard deviation s estimate the common population voxel mean μ and standard deviation σ. Regardless of a normality assumption, Var $(X_C-\bar{X})=\sigma^2(1+\frac{1}{N})$ and Var $(X_R-\bar{X})=\sigma^2(1-\frac{1}{N})$, where N is the reference group sample size and Var refers to variance. These variances are different because X_C is statistically independent of $\bar{X}$, whereas X_R and $\bar{X}$ are positively correlated due to $\bar{X}$ being computed from a sample that includes X_R. Incorrectly accounting for these differences in variance may lead to bias. Consequently, different standardizations of $(X_C-\bar{X})$ and $(X_R-\bar{X})$ are warranted, suggesting that the distributions of Z_C and Z_R are likely different.

Geisser notes that $T_c=\frac{X_C-\bar{X}}{\left\{s\sqrt{1+\frac{1}{N}}\right\}}=Z_c\sqrt{\frac{N}{(N+1)}}$ has a Student's t distribution with $N-1$ degrees of freedom, written symbolically as $T_C \sim t_{N-1}$ (Geisser, 1993). Cook and Weisberg show that the distribution of the studentized residual $T_R=\frac{X_R-\bar{X}}{\left\{s\sqrt{1-\frac{1}{N}}\right\}}=Z_R\sqrt{\frac{N}{(N-1)}}$ is symmetric about zero with a standard deviation of 1, and that $\frac{T_R^2}{N-1}$ has a Beta distribution with parameters 0.5 and $0.5(N-2)$ (Cook and Weisberg, 1982). As a result, the distributions of Z_C and Z_R differ, which leads to different probabilities that Z_C and Z_R are more extreme at a common fixed threshold, even when the underlying distributions of X_C and X_R are identical.

Our proposed DisCo-Z correction modifies the z-thresholds so that exceedance probabilities for the two distributions are identical in the null case. We first specify a fixed threshold probability (e.g., $\alpha < 0.05$), and desired upper tail thresholds c_N and r_N that are necessarily a function of N so that

$$\alpha=\Pr(Z_R > r_N)=\Pr(Z_C > c_N) \quad (1)$$

The arbitrary choice of $\alpha = 0.0228$ corresponds to the probability a standard normal random variable exceeds a z-score of 2. Using the distributions given above, we demonstrate that the adjusted value for the comparison group is given by

$$c_N=t_{1-\alpha,N-1}\sqrt{1+\frac{1}{N}} \quad (2)$$

where $t_{1-\alpha,\ N-1}$ is the $100(1-\alpha)$ percentile of the t_{N-1} distribution and N is the reference group sample size. It is also notable that Eq. (2) corresponds to the one-versus-many test that has been previously suggested for examining a single subject's data against a reference population (Lipton et al., 2008; Patel et al., 2007), as well as the primary transformation that occurs when using the leave-one-out method. Importantly, the shape of the t-distribution is dependent on N, especially at lower sample sizes. Thus, equivalent z-thresholds are not necessarily applicable for the reference and comparison groups using the leave-one-out method in the case of unequal N between reference and comparison groups.

In contrast, the adjusted threshold for the z-transformed values for the reference group is given by

$$r_N=(N-1)\sqrt{\frac{B(1-2\alpha,0.5,0.5[N-2])}{N}} \quad (3)$$

where $B(1-2\alpha, 0.5, 0.5[N-2])$ is the $100(1-2\alpha)$ percentile of a Beta distribution with parameters 0.5 and $0.5(N-2)$. Z-transforming longitudinal (i.e., correlated) data from the reference group is dependent on the degree of covariance between observations on the same individual. A derivation for transforming longitudinal data from the reference group, and the fundamental statistical relationship between the DisCo-Z and leave-one-out approach, is presented in supplementary materials.

The DisCo-Z thresholds are therefore different for the two samples and dependent on N, with the degree of adjustment decreasing as a function of increasing N (Supporting Information Table 1). Although these effects will be present during the z-transformation of any identical normal distributions (e.g., clinical data), appropriate corrections are more critical in neuroimaging studies given the large number of voxels that are z-transformed.

Monte Carlo simulations (simulated and real data) are first used to compare performance of the pothole, independent reference sample and DisCo-Z methods for identifying extreme values in simulated and DTI data derived from 50 HC, where no differences between randomly selected samples were expected. The second aim was to evaluate the robustness of the DisCo-Z method on simulated data with different distribution properties (normal, t, and chi-square) that may more closely represent the data acquired with most neuroimaging techniques, and to evaluate other commonly used methods. Finally, previously published data from our lab (Ling et al., 2012; Mayer et al., 2012) that compared SSA between mTBI patients and control populations using the pothole method were also re-examined.

METHODS AND RESULTS

The current investigation was conducted on two DTI datasets presented in previous publications. The first dataset (Ling et al., 2012) included 50 adult mTBI patients (25 males; 27.88 ± 9.22 years old; 13.12 ± 2.21 years of education) and 50 matched adult HC (25 males; 27.42 ± 8.96 years old; 13.90 ± 2.09 years of education). The adult HC dataset was used for all Monte Carlo simulations described below. The second dataset (Mayer et al., 2012)

included 15 pediatric mTBI patients (13 males; 13.47 ± 2.20 years old; 6.87 ± 2.23 years of education) and 15 matched children (12 males; 13.40 ± 1.84 years old; 7.27 ± 1.87 years of education), and was used for the purpose of data re-analyses only. Participants were excluded from the study if there was a history of neurological disease, psychiatric disturbance, additional closed head injuries with more than 5 minutes loss of consciousness, a head injury within the last year, learning disorder, ADHD, or a history of recent substance or alcohol abuse. Informed consent was obtained from all participants according to institutional guidelines at the University of New Mexico.

MR Imaging

All images were collected on a 3 Tesla Siemens Trio scanner. Foam padding and paper tape were used to restrict motion within the scanner. High resolution T1-weighted anatomic images were acquired with a 5-echo multi-echo MPRAGE sequence (TE = 1.64, 3.5, 5.36, 7.22, 9.08 ms, TR = 2.53 s, TI = 1.2 s, 7° flip angle, NEX = 1, slice thickness = 1 mm, FOV = 256 × 256 mm, voxel resolution = 1 × 1 × 1 mm^3). A single run (pediatric sample) or two runs (adult sample) of DTI scans ($b = 800$ s/mm^2) were acquired using a twice-refocused spin echo sequence with 30 diffusion gradients and the b = 0 experiment repeated five times (72 interleaved slices; TE = 84 ms; TR = 9 s; 90° flip angle; NEX = 1; slice thickness = 2.0 mm; FOV = 256 × 256 mm; matrix size = 128 × 128; voxel resolution = 2 × 2 × 2 mm^3). GRAPPA (2× acceleration and 32 reference lines) was used to reduce susceptibility-induced image distortions.

The AFNI software package (Cox, 1996) was used to process and analyze DTI datasets. In the adult samples, the raw DTI data and gradient tables were first concatenated across the two runs. Image distortions caused by eddy currents and head motion from both samples were next corrected by registering all diffusion weighted images to the first $b = 0$ s/mm^2 image using a 12 degree of freedom (df) affine correction with mutual information as the cost function. The vector corresponding to the rotation component was then extracted from the resultant transformation matrix and applied to the gradient table. Prior to calculating diffusion tensors and scalar measures (fractional anisotropy [FA]), images were smoothed anisotropically to improve signal to noise characteristics (Ding et al., 2005). A non-linear method was adopted for tensor calculations to decrease tensor estimate errors caused by noise, especially in regions of high anisotropy (Cox and Glen, 2006). Diffusion weighted images were registered to the subject's T1 anatomic image using an affine transformation with 12 df and Local Pearson Correlation as the cost function (Saad et al., 2009). This transformation matrix was then multiplied by the matrix corresponding to T1 stereotaxic normalization such that each participant's FA data was normalized to Talairach space. Each subject's FA data was then blurred with a 6 mm FWHM kernel. To reduce the number of comparisons and restrict the analysis to white matter, all data was masked by the Johns Hopkins University (JHU) white matter labels atlas from FSL (Mori and van Zijl, 2007).

Standard Pothole Approach

Analyses using the Shapiro–Wilk test were first conducted to determine the degree of deviation from normality for the HC data ($N = 50$) within the JHU mask. Results indicated that 39,635 (26.92%) of the 147,244 voxels were non-normally distributed ($P < 0.05$). Of the non-normal voxels, 7,681 (19.38%) showed evidence of a negative skew (< -0.5) as measured using the third moment about the mean divided by the standard deviation (Bulmer, 1979).

The next series of analyses evaluated the pothole method (White et al., 2009). First, the spatially normalized whole-brain FA maps from 50 HC were randomly sampled with replacement into either a reference (RF) or comparison (CP) group to maintain statistical independence in the sampling procedure. Sample sizes of $N = 10$ to $N = 50$ per group were evaluated, using steps of $N = 5$. Second, the mean and standard deviation of FA were calculated for each voxel from the reference group. Third, individual subject data for both the reference and comparison groups were transformed to signed z-scores on a voxel-wise basis using the statistical moments derived from the reference group. Extreme voxels in both groups were identified based on two standard deviations above ($z > 2$; positive) or below ($z < -2$; negative) the reference's voxel mean. Fourth, a minimum cluster size threshold of 128 μL (16 native voxels) was also applied to the data to reduce the likelihood of false positives (Kim et al., 2013; White et al., 2009).

Fifth, the number of voxels exceeding both the magnitude and spatial cluster threshold were then summed for each subject, as were the number of surviving clusters. There were a total of 147,244 1-mm isotropic voxels in the JHU atlas, suggesting that approximately 3,358 voxels (2.28%) should survive the magnitude threshold of $z = |2|$ per tail assuming a normal distribution with spatially independent voxels. However, this value was only used as a general benchmark given the native smoothness of MRI data and the 6 mm blur that was applied to the data. Finally, two-tailed independent samples t-tests compared whether the number of voxels and clusters that exceeded the z-threshold were statistically different between the reference and comparison groups ($P < 0.05$). Statistically significant t-tests were summarized to reflect whether the mean number of extreme voxels was greater for the RF (i.e., RF > CP) or the CP group (i.e., CP > RF) for both positive and negative tails. The entire procedure (steps 1–6) was then iterated 100 times. The magnitude of bias was quantified using the percentage of t-tests that were significantly different across groups. In the absence of bias and

under the null distribution, approximately 5% of the *t*-tests comparing the reference and comparison groups should have been significant and equally distributed (CP > RF ≈ RF > CP).

As predicted by Eqs. (2) and (3), results of these simulations revealed a significant bias for the comparison group (CP > RF) when using the pothole method. Importantly, the number of positive and negative extremes (voxels or clusters) exceeded expectations (3,358 voxels) for the comparison group and was below expectations for the reference group (Fig. 1A; rows 1 and 3), with significant bias present for comparison relative to reference group (CP > RF) in both the positive and the negative tails (Fig. 1A; rows 2 and 4). The magnitude of bias varied as a function of sample size. At $N = 10$ the comparison group exhibited a statistically significantly greater mean number of extreme voxels (positive extremes = 50% of iterations; negative extremes = 69%; Fig. 1A, row 2) and clusters (positive extremes = 68%; negative extremes = 68%; Fig. 1A, row 4) relative to the reference group. Although the bias decreased rapidly with increasing sample size, significant bias for the comparison group was still present for voxels (positive extremes = 25%; negative extremes = 28%) as well as clusters (positive extremes = 31%; negative extremes = 26%) at sample sizes ($N = 30$) that are typically equivalent or greater than most published DTI studies. Finally, the number of extreme voxels was typically larger for negative relative to positive tails across all sample sizes. Supplemental analyses indicated that results were similar when non-parametric tests were used for group comparisons to estimate bias.

Bias Correction Using the DisCo-Z Method

Next, we evaluated whether our theoretically determined adjustments for z-thresholds would be sufficient for correcting the biases that existed in both the reference and comparison groups for the pothole method. The samples and methodology used in these analyses were identical to the pothole method with the exception that instead of applying a uniform z-threshold (e.g., $z > 2$; $z < -2$) to the data from both groups, adjusted z-thresholds were now determined by either Eq. (2) (comparison group) or (3) (reference group). Independent samples t-tests were again used to assess the presence of bias in the z-transformed data between the reference and comparison groups (uncorrected value of $P < 0.05$).

Results indicated that the DisCo-Z method eliminated almost all bias regardless of sample size. Specifically, the percentage of significant differences between the two comparison groups approximated alpha when summed across the group *t*-tests (Fig. 1B, rows 2 and 4). There was no evidence of bias even when *P*-value was increased to 0.10 (see Supporting Information Fig. 1). Moreover, in contrast to the pothole method, the number of surviving voxels (Fig. 1B, row 1) and clusters (Fig. 1B, row 3) was consistent across all sample sizes, with the number of voxels approaching theoretical predictions (3,358 voxels) for negative but not positive extremes. Similar to the pothole results, the number of extreme negative voxels was again higher than the number of positive voxels, which likely resulted from the predominately negative skew in a minority (≈19%) of the voxels.

Standardization with an Additional Reference Sample

Other approaches for correcting bias include variations on the use of statistically independent groups to generate the statistical moments, followed by z-transformations of the data (Bouix et al., 2013; Kim et al., 2013). To evaluate this approach, the 50 HC were first randomly assigned with replacement to either a reference or a comparison group of size *N*. Participants in the comparison group were then randomly sampled with replacement to achieve two separate comparison groups (CP1 and CP2) of similar *N*. Otherwise identical procedures were applied to the data as used with the pothole and DisCo-Z method. Independent t-tests were then conducted to compare both voxel counts and the number of significant clusters between the two comparison groups on the z-transformed data.

Results indicated that the number of surviving voxels/clusters were inversely proportional to sample size for both comparison groups given the properties of a t-distribution at smaller *N* (see Fig. 1C). However, there was no evidence of group-wise bias between the two comparison groups across different sample sizes for either the voxel count data or the number of clusters (significant differences between comparison groups approximately 5%). Negative extremes were again always significantly larger than positive extremes.

Simulations with Normal and Non-Normal Distributions

The theoretical basis for the DisCo-Z method described in the introduction assumes a normal distribution. However, this assumption was violated in almost 27% of the voxels in our DTI dataset, and an assumption of normality may not be true for data derived from different imaging modalities. Therefore, Matlab was used to simulate data derived from normal, *t*, and chi-square distributions to further evaluate the robustness of the DisCo-Z method as well as previously proposed methods. To maintain consistency with previous analyses, the same number of data points (e.g., 147,244 "voxels") was used in each of the simulations. Similar to our DTI results, the simulations were run for sample sizes 10–50 for the reference and comparison groups in steps of 5, with the number of iterations increased to 400.

In the first simulation, values for the reference and comparison groups were randomly sampled from a standard

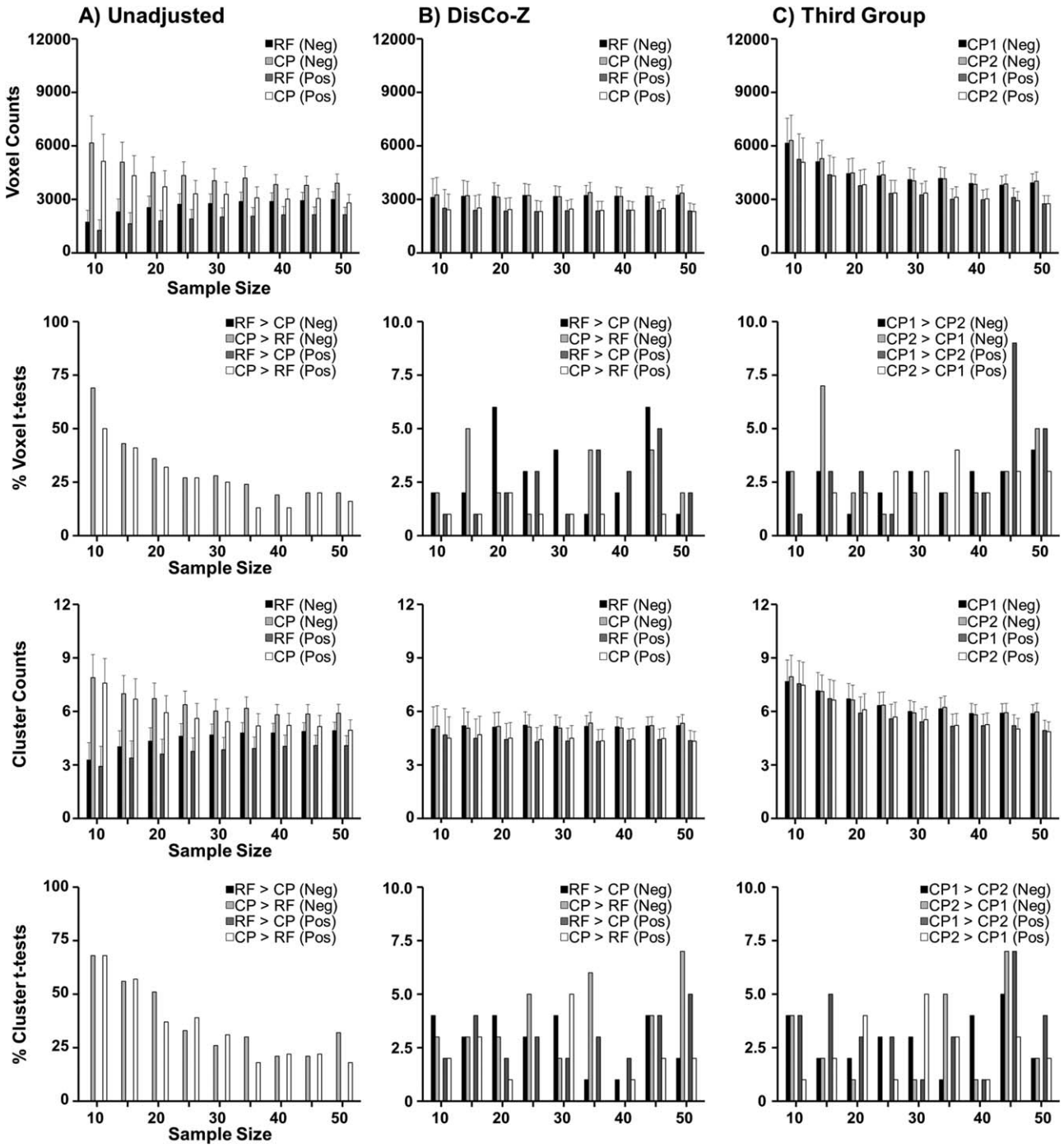

Figure 1.

This figure presents results of z-transformed FA data derived from 50 HC at sample sizes from 10 to 50 in increments of 5 using the pothole method (Column A), DisCo-Z method (Column B), or using an independent sample (Column C). For all data, the first row presents the number of extreme voxels that exceeded statistical (Columns A and C = $|z| > |2|$; Column B DisCo-Z determined threshold) and spatial (16 native voxels) thresholds, whereas the second row depicts the percentage of *t*-tests that were statistically different ($P < 0.05$) between the groups. The third row presents the number of clusters surviving the positive and negative thresholds, with the fourth row depicting the percentage of significant *t*-tests. Error bars on the first and third rows depict the average standard error of the mean across iterations. Note that the scaling for the *t*-test graphs (second and fourth row) is different for the pothole method (Column A; maximum 100) relative to the two correction methods (Columns B and C; maximum 10). Columns A and B are color-coded to reflect results from both the negative (Neg) and positive (Pos) tails of the distribution for the reference (RF Neg = black bars; RF Pos = dark gray bars) and comparison (CP Neg = light gray bars; CP Pos = white bars) groups. Results from the *t*-tests are similarly color-coded to indicate whether bias in the estimation of extreme values was more prevalent in the CP relative to RF group for each of the tails (e.g., CP > RF (Neg) = light gray). Column C also presents results from two comparison groups (CP1 and CP2) when the statistical moments were derived from a third independent group that served as the reference. The series of graphs clearly depict the over-estimation (CP) and under-estimation (RF) of z-scores and resultant bias when using the pothole method, and the elimination of the bias with either the DisCo-Z method or through an independent group.

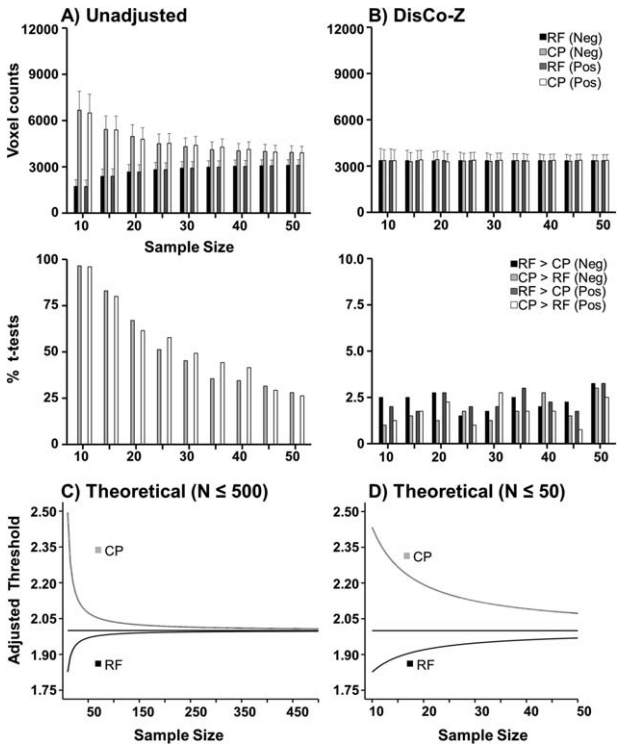

Figure 2.

Data for these graphs were sampled from a standard normal distribution ($\mu = 0$, $\sigma = 1$) at sample sizes between 10 and 50 in increments of 5. Row 1 looks at extreme values (Panel A: $|z| > |2|$ for pothole method; Panel B: DisCo-Z adjusted threshold) in the negative (Neg) and positive (Pos) tails of the distribution. The number of data points surviving the statistical threshold for both the reference (RF Neg = black bars; RF Pos- = dark gray bars) and comparison (CP Neg = light gray bars; CP Pos = white bars) groups are presented in the first row, whereas the second row presents the percentage of significant *t*-tests ($P < 0.05$) using a similar color scheme to indicate whether bias was greater in the CP relative to RF group. Voxel count error bars depict the average standard error of the mean across iterations. Note that the scaling for *t*-test graphs is different for the unadjusted threshold (maximum 100) relative to the DisCo-Z method (maximum 10). Panels C and D plot statistically derived z-thresholds for eliminating bias in the comparison group [CP; light gray line derived from Eq. (2)] and reference group [RF; black line derived from Eq. (3)] and against a common study specific z-threshold ($z = 2$, horizontal line) as a function of sample size. Panel C depicts z-thresholds for sample sizes ranging from 10 to 500, while Panel D focuses in on sample sizes from 10 to 50.

normal distribution ($\mu = 0$, $\sigma = 1$). For the second simulation, a *t*-distribution (6 df) was selected to examine the effects of kurtosis (i.e., sub-Gaussian distribution with heavy tails). The third and fourth simulations randomly sampled values for the reference and comparison groups from two chi-squared distributions with different degrees of freedom to model a stronger (6 df) or weaker (12 df) skew. The chi-squared and t-distributions were also scaled to have a variance of 1. For all simulations, the data were further constrained to have an average interclass-correlation of 0.10, which roughly corresponded to the correlation measured from 500 randomly selected voxels in our HC sample (mean $r = 0.12$; sd = 0.20). Examples of the simulated distributions are presented in Supporting Information Figure 2, as well as the effects of z-transforms using the pothole approach on both the reference and comparison groups.

For each of the iterations, a reference and a comparison group of a given sample size was randomly generated from the four different distributions. The data were then z-transformed using the statistical moments from the reference group. Extreme voxels were identified in both the reference and comparison groups based on either identical thresholds ($z > 2$; $z < -2$) or by the DisCo-Z method. Group-wise tests (independent *t*-tests) were then conducted between the comparison and reference groups on the number of surviving voxels for both the unadjusted and adjusted methods. The number of significant *t*-tests ($P < 0.05$) was then computed across all iterations.

In general, the results from the simulated, normally-distributed data were similar to unadjusted and DisCo-Z results from the DTI data (Fig. 2A, B), and closely followed theoretical predictions (Fig. 2C, D). The largest differences between the reference and comparison groups were present at the smaller ($N = 10$) sample sizes (positive extremes = 96%; negative extremes = 96.5%), with the degree of bias diminishing as a function of increasing sample ($N = 30$; positive extremes = 49.25%; negative extremes = 45.25%). Appropriate statistical adjustment of the threshold values using the DisCo-Z method eliminated the bias across all sample sizes. The results from the *t*-distribution (Fig. 3) were similar in terms of magnitude of bias ($N = 10$: positive extremes = 94.75%; negative extremes = 94.5%) as well as the elimination of bias with the DisCo-Z method.

Similar results were also observed for the chi-squared simulation using either 6 ($N = 10$: positive extremes = 97.2%; negative extremes = 53.5%) or 12 ($N = 10$: positive extremes = 97.8%; negative extremes = 77.3%) df (Figs. 4 and 5) distributions. As evidenced by the chi-square results, the degree of bias between the reference and comparison groups for positive extremes was also affected by the degree of skew within the distribution (bias in Fig. 4A < Fig. 5A). In contrast, bias for negative extremes remained relatively constant regardless of the degree of skew. This was confirmed by reversing the direction of skew by multiplying values sampled from the initial distribution by -1. In negatively skewed distributions, the bias now changed as a function of skew only for negative extremes (bias in Supporting Information Figs. 3A < 4A).

Finally, additional simulations were performed to test the effects of using an independent sample as a reference

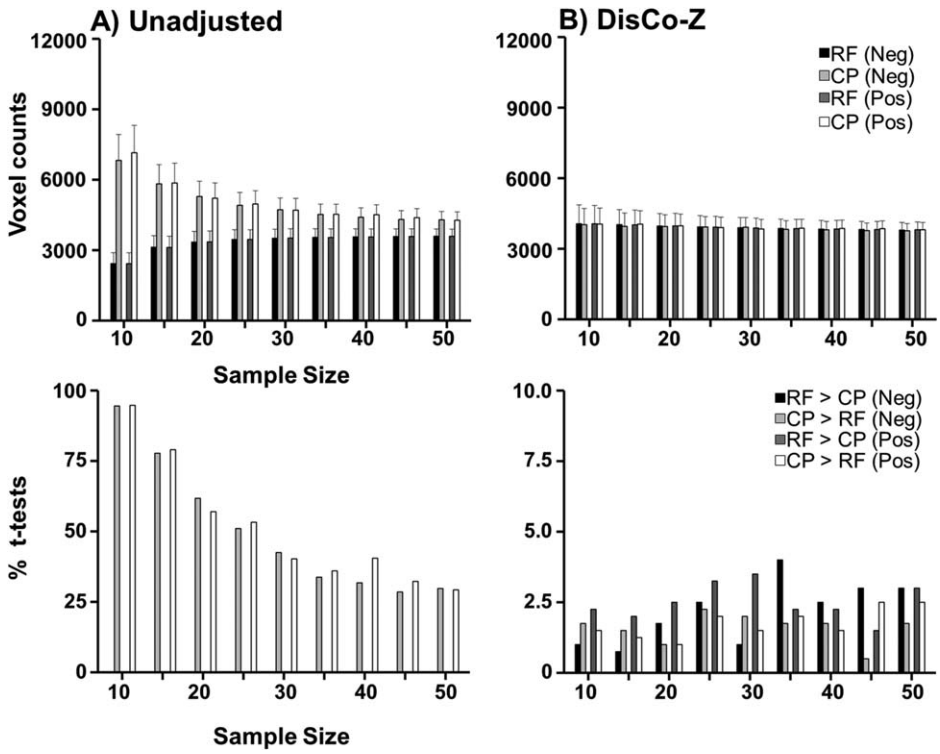

Figure 3.

Data for these graphs were sampled from a population with a *t*-distribution (6 df) at sample sizes between 10 and 50 in increments of 5. Row 1 looks at extreme values (Panel A: $|z| > |2|$ for pothole method; Panel B: DisCo-Z adjusted threshold) in the negative (Neg) and positive (Pos) tails of the distribution. The number of data points surviving the statistical threshold for both the reference (RF Neg = black bars; RF Pos = dark gray bars) and comparison (CP Neg = light gray bars; CP Pos = white bars) groups are presented in the first row, whereas the second row presents the percentage of significant *t*-tests ($P < 0.05$) using a similar color scheme to indicate whether bias was greater in the CP relative to RF group. Voxel count error bars depict the average standard error of the mean across iterations. Note that the scaling for *t*-test graphs is different for the unadjusted threshold (maximum 100) relative to the DisCo-Z method (maximum 10).

group (Fig. 6A), the leave-one-out method (Fig. 6B), and the EZ-MAP approach (Fig. 6C). The normal distribution was chosen for these additional simulations, and identical methods were utilized as in Figure 2. Similar to DTI results presented in Figure 1C, the use of an independent third reference group with two comparison groups eliminated all evidence of group-wise bias between comparison groups. The resulting distributions from both CP1 and CP2 generally followed the predicted *t*-distribution as evidenced by the decrease in extrema as a function of *N* (Fig. 6A). For the leave-one-out method, each observation from the reference group was z-transformed based on the remainder of observations, whereas the comparison group was z-transformed based on the entire reference group. Identical z-thresholds were then applied to both groups, as has been commonly implemented in the literature. As expected, results from the leave-one-out method were similar to the independent group with one important exception (Fig. 6B). Specifically, as predicted by statistical theory, there was a small but consistent evidence of bias between the RF and CP groups (RF > CP) at smaller sample sizes due to the differential *N* used to z-transform each group (RF = $N - 1$; CP = N). This is a direct result of using identical z-thresholds for both groups, which affects the probability of extrema more in small sample sizes due to the shape of *t*-distributions as a function of *N*.

The EZ-MAP method was also implemented with an independent sample serving as the reference group based on published reports. Specifically, the data from the two comparison groups was first transformed using the mean and standard deviation from the reference sample, and then divided by a bootstrapped standard deviation derived from the reference sample to account for variance under-estimation (Kim et al., 2013). In addition to eliminating group bias (Fig. 6C), the EZ-MAP method also reduced the number of extrema observed in both comparison groups to statistically predicted levels (~3,358 extrema per tail). Importantly, as evidenced by Supporting

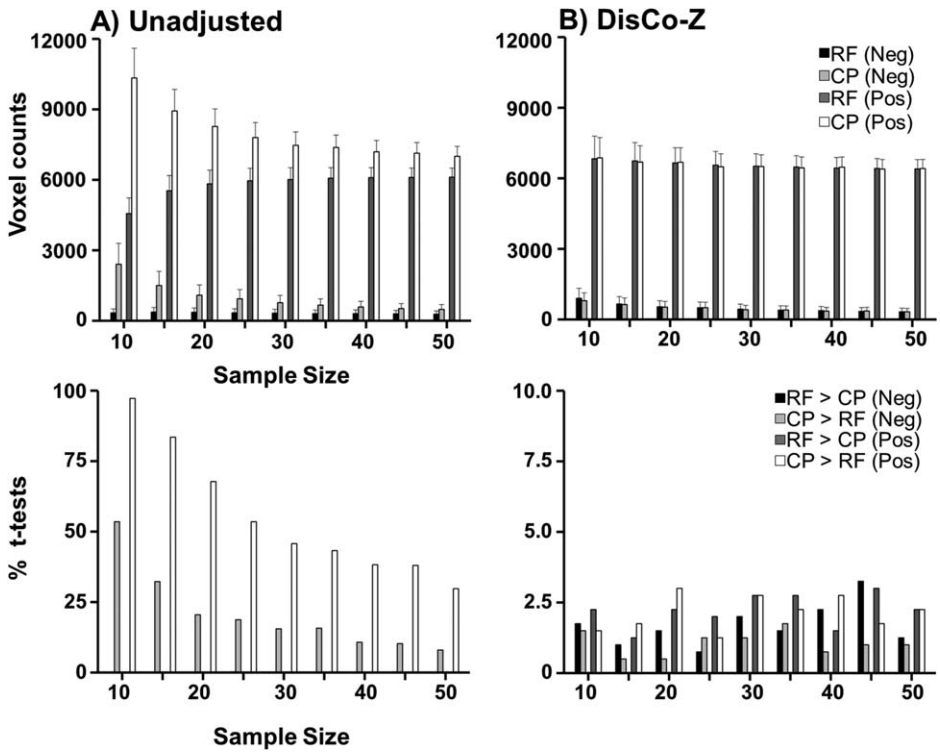

Figure 4.

Data for these graphs were sampled from a highly positively skewed chi-squared distribution (6 df) at sample sizes between 10 and 50 in increments of 5. Row 1 looks at extreme values (Panel A: $|z| > |2|$ for pothole method; Panel B: DisCo-Z adjusted threshold) in the negative (Neg) and positive (Pos) tails of the distribution. The number of data points surviving the statistical threshold for both the reference (RF Neg = black bars; RF Pos = dark gray bars) and comparison (CP Neg = light gray bars; CP Pos = white bars) groups are presented in the first row, whereas the second row presents the percentage of significant *t*-tests ($P < 0.05$) using a similar color scheme to indicate whether bias was greater in the CP relative to RF group. Voxel count error bars depict the average standard error of the mean across iterations. The evidence of bias in the pothole method was eliminated through statistical adjustment of z-thresholds despite the significant skew of the sample, although significant differences emerged in the number of negative and the positive extremes as well as the subsequent bias in each of the tails. Note that the scaling for *t*-test graphs is different for the unadjusted threshold (maximum 100) relative to the DisCo-Z method (maximum 10).

Information Figure 5, using the EZ-MAP method with a single reference sample resulted in group-wise bias (RF > CP) that decreased as a function of sample size. Specifically, in this simulation the bias associated with the comparison group is corrected whereas the bias with the reference group is not [Eq. (3)].

Re-Analyses of Previous Data

Finally, we applied the DisCo-Z correction method to two previously published studies in which the pothole method was utilized (Ling et al., 2012; Mayer et al., 2012). The first study reported increased FA within the genu of the corpus callosum and several other ROI in 50 semi-acutely (i.e., <21 days post) injured mTBI patients relative to 50 HC (Ling et al., 2012). Result from the pothole analysis indicated a greater number of clusters with increased FA ($F_{1,97} = 6.41$, $P = 0.013$, Cohen's $d = 0.54$) for mTBI patients relative to HC during the semi-acute injury phase, with no group differences observed for clusters with decreased FA ($P > 0.10$). Twenty-six adult mTBI patients and 26 HC returned for a follow-up visit approximately 4 months post-injury. Results from a longitudinal pothole 2 × 2 (Group × Time) mixed measures ANCOVA analysis indicated a trend in the Group × Time interaction ($F_{1,49} = 3.68$, $P = 0.061$) for the total number of positive clusters. Simple-effects testing indicated that clusters with increased FA were significantly reduced at Visit 2 for the mTBI patients ($t_{1,25} = 2.40$, $P = 0.024$) but were unchanged for HC ($P > 0.10$).

The results from the pothole analyses were therefore repeated using identical methods, with the exception of

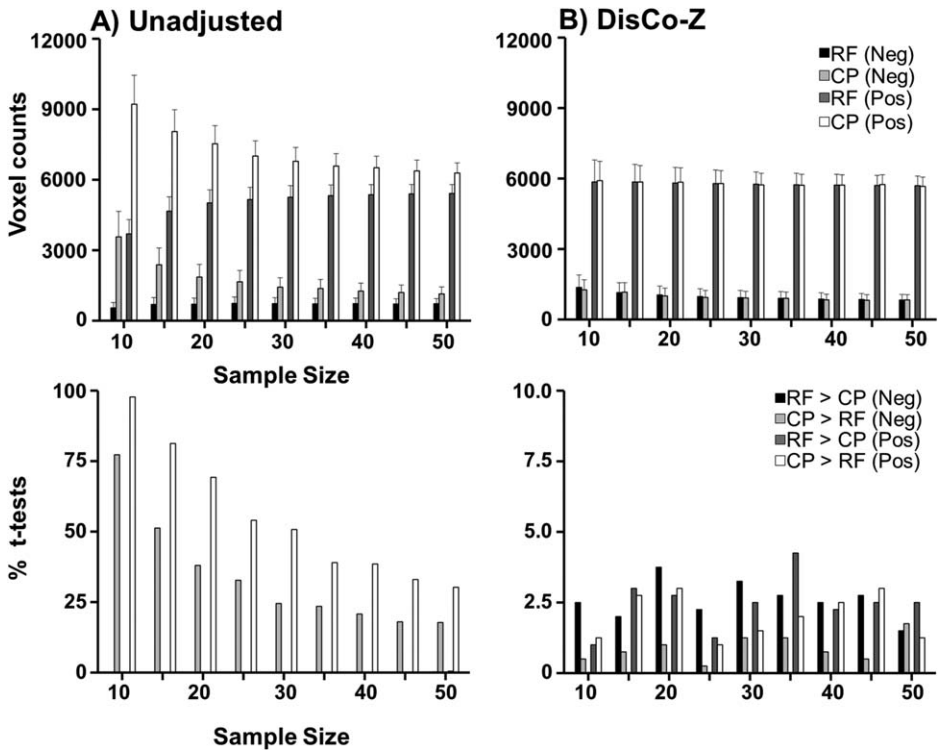

Figure 5.

Data for these graphs were sampled from a moderately positively skewed chi-squared distribution (12 df) at sample sizes between 10 and 50 in increments of 5. Row 1 looks at extreme values (Panel A: $|z| > |2|$ for pothole method; Panel B: DisCo-Z adjusted threshold) in the negative (Neg) and positive (Pos) tails of the distribution. The number of data points surviving the statistical threshold for both the reference (RF Neg = black bars; RF Pos = dark gray bars) and comparison (CP Neg = light gray bars; CP Pos = white bars) groups are presented in the first row, whereas the second row presents the percentage of significant *t*-tests ($P < 0.05$) using a similar color scheme to indicate whether bias was greater in the CP relative to RF group. Voxel count error bars depict the average standard error of the mean across iterations. Note the elimination of bias with the DisCo-Z method, with differential effects in the negative and positive tails of the distribution. The scaling for *t*-test graphs is different for the unadjusted threshold (maximum 100) relative to the DisCo-Z method (maximum 10).

the adoption of the recommended DisCo-Z thresholds. Reanalyzed results indicated that number of clusters with increased FA were no longer statistically significant ($F_{1,97} = 1.86$, $P = 0.175$, $d = 0.29$) between mTBI patients and controls at Visit 1. Cluster metrics of decreased FA remained non-significant ($P > 0.10$) as well. Visit 2 data were also reanalyzed to assess for dynamic change in the mTBI group. Because of the expected correlation between Visit 1 and Visit 2 data, and the derivation of the statistical moments used in z-transforms from HC (i.e., reference sample) Visit 1 sample, the resulting Visit 2 z-transformed distribution for HC should be intermediate to a scaled signed square root of a Beta (perfect correlation between Visits 1 and 2) and a *t* (no correlation between Visits 1 and Visit 2) distribution (see Supporting Information Materials). In contrast, the patient data at both time-points is statistically independent from the Visit 1 HC data, such that a *t*-distribution [Eq. (2)] can still be applied to correct the Visit 2 mTBI data. Results from the longitudinal analyses (paired *t*-tests) using the DisCo-Z correction method indicated a reduction in the number ($t_{1,25} = 3.59$, $P = 0.001$) and volume ($t_{1,25} = 3.81$, $P = 0.001$) of clusters with increased anisotropy at Visit 2 for the mTBI patients. Reanalysis of the Visit 2 data for clusters metrics regarding decreased FA remained non-significant ($P > 0.10$).

Our second study examined diffusion abnormalities in a smaller cohort of 15 pediatric mTBI patients and 15 pediatric HC (Mayer et al., 2012). ROI results again indicated increased FA within the right and left anterior corona radiata, and left cerebral peduncles ($P < 0.05$), with a non-significant trend for the left superior corona radiata ($P = 0.052$) for patients relative to controls. A standard voxel-wise analysis indicated increased FA for pediatric mTBI patients within several white matter tracts following appropriate corrections for false positives ($P < 0.05$). Pothole analyses indicated significantly increased number

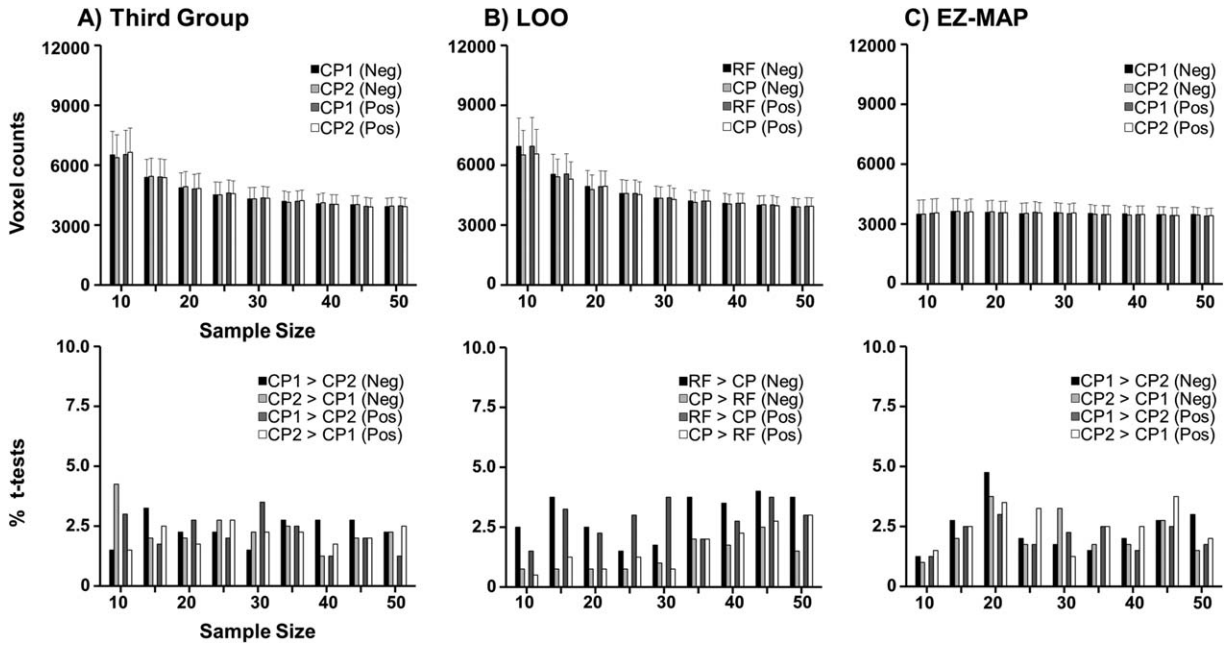

Figure 6.

This figure presents simulated data derived from a normal distribution at sample sizes ranging from 10 to 50 in increments of 5. Methods of comparison include using a third reference sample to derive statistical moments for two comparison groups (CP1 and CP2; Column A), the leave-on-out method (Column B), or the EZ-MAP method (Column C). The first row presents the number of extreme voxels that exceeded statistical thresholds as determined by each model, whereas the second row depicts the percentage of *t*-tests that were statistically different ($P < 0.05$) between the groups. Error bars on the first row depict the average standard error of the mean across iterations. Columns A and C are color-coded to reflect results from both the negative (Neg) and positive (Pos) tails of the distribution for two comparison groups derived from a third independent sample (CP1 Neg = black bars; CP1 Pos = dark gray bars; CP2 Neg = light gray bars; CP2 Pos = white bars). Column B is color-coded to reflect results from both the negative (Neg) and positive (Pos) tails of the distribution for the reference (RF Neg = black bars; RF Pos = dark gray bars) and comparison (CP Neg = light gray bars; CP Pos = white bars) groups. Results from the *t*-tests are similarly color-coded to indicate whether bias in the estimation of extreme values was more prevalent in the CP relative to the RF group for each of the tails (e.g., CP > RF (Neg) = light gray). The series of graphs show that all methods reduce group-wise bias. The EZ-MAP method reduces the number of extrema to statistically predicted levels at all *N*, while the number of extrema through the use of independent samples decreases as a function of *N*.

($t_{28} = 7.27$, $P < 0.00001$, $d = 2.7$) and volume ($t_{28} = 6.98$, $P < 0.00001$, $d = 2.6$) of clusters with increased FA for pediatric mTBI patients relative to HC at Visit 1, with no significant differences for cluster metrics of decreased FA. A 2 × 2 (Group × Visit) longitudinal pothole analyses in 10 returning pediatric mTBI subjects and matched controls indicated a significant ($F_{2,17} = 15.67$, $P < 0.0001$) multivariate effect of group for the number and volume of clusters with increased FA, with a non-significant Group × Visit interaction.

Reanalyses of Visit 1 data using the DisCo-Z method indicated that the group difference in the number ($t_{28} = 2.29$, $P = 0.030$, $d = 0.84$) and volume ($t_{28} = 2.57$, $P = 0.016$, $d = 0.97$) of clusters with increased FA remained significantly increased for the pediatric mTBI sample. In addition, HC also exhibited a greater number ($t_{21} = -2.95$, $P = 0.008$) and volume ($t_{22.3} = -3.23$, $P = 0.004$) of clusters with decreased FA. However, the latter finding resulted from a 0 cluster count in 9/15 of the mTBI patients. Reanalysis of the longitudinal data with DisCo-Z method indicated that there was no significant change ($P > 0.10$) in the number or volume of clusters with either increased or decreased FA for the pmTBI patients at 4 months post-injury.

DISCUSSION

Several analytic techniques have been proposed for capturing regions of abnormal signal ("lesions") on a subject-specific basis, including variations on normative (i.e., *z*-scores) transformations (Bouix et al., 2013; Davenport et al., 2012; Jorge et al., 2012; Ling et al., 2012; Mayer et al., 2012; White et al., 2009), cross-validation (Viviani et al.,

2007), bootstrapping (Bazarian et al., 2012), as well as a combination of methods (Kim et al., 2013; Lipton et al., 2012). Results from the current study indicate that z-transforming individual participant data based on the statistical moments from the entire reference group alone (pothole method) resulted in bi-directional bias for both the reference (under-estimated the number of extremes) as well as comparison (over-estimated the number of extremes) groups. As suggested by statistical theory, this is a direct result of the different distributions that result from z-transforming both the reference (normal to scaled signed square root of a Beta distribution) and comparison (normal to *t*-distribution) samples (see Supporting Information Fig. 2) in part due to incorrect variance estimation.

Thus, the pothole method is likely to return a different number of statistically determined extremes even when samples are derived from the same population. In the current study, the number of surviving voxels in the reference and comparison groups following z-thresholding converged with increasing N but did not reach statistical equivalence (i.e., no group differences) even at sample sizes of 50. Thus, bias with the pothole method is likely to be present at sample sizes typically used in most neuroimaging studies ($N = 30$ per group) and even beyond. This is likely a result of the large number of voxels that must be z-transformed in voxel-wise analyses, increasing the likelihood of bias even when the theoretical differences in the underlying z-thresholds of the transformed data are relatively small due to large N (Fig. 2C).

Several different methods were evaluated for eliminating this bias in the current study. One set of methods involved the utilization of an "independent" reference group, which can include a single reference group (e.g., typically "healthy") and two separate comparison (e.g., one "healthy" and one "patient") groups, or a single reference group in which the individual data-points are separately z-transformed (leave-one-out method). The principal disadvantage of a third independent sample is the costly nature of MR-data acquisition and the potential allure of using convenience samples (i.e., existing data from young, healthy right-handed individuals) to form the initial reference group. The leave-one-out approach eliminates the costs associated with data acquisition and processing of a separate reference group. However, the leave-one-out approach as currently implemented is computationally more difficult given that the number of different z-transformations is essentially equivalent to the sample size of the reference group (but see Supplemental Methods for a statistical derivation). It can also lead to bias when z-thresholds are not correctly adjusted for differing degrees of freedom [Eq. (2)] between the reference and comparison groups (typically $N - 1$) in studies with very small N. Finally, both the independent reference sample and the leave-one-out approach resulted in over-estimation of extremes for both comparison groups rather than identifying the desired number of extremes based on probability theory (e.g., 2% of extreme data points). Importantly, over-estimation of extremes in conjunction with any independent sample method (a third sample or leave-one-out) can easily be corrected by using Eq. (2).

The EZ-MAP method has also been proposed for comparing two groups in conjunction with a third independent reference group (Kim et al., 2013; Lipton et al., 2012). The EZ-MAP method improves the scaling of the *z*-scores by estimation of the bootstrapped variance, which approximates unity as a function of increasing sample size (Kim et al., 2013). As noted above, the utilization of an independent reference group (either an entire group or through the leave-one-out methodology) will eliminate the primary bias in z-transformed data from two comparison samples due to distributional equivalence (i.e., two *t*-distributions). The EZ-MAP method more closely resembles a standard normal z-distribution by correcting for inaccuracies in variance estimation (deviation from unity) through repetitive sampling of the reference group. However, bootstrapping approaches are also computationally expensive. Importantly, the EZ-MAP approach should not be applied to two-sample data (i.e., one patient and one control group) when trying to estimate whether the number of extrema vary between groups. This approach would result in bias due to distributional changes to the reference group following the z-transformation [Eq. (3)] in conjunction with effective correction for the comparison group (Supporting Information Fig. 5).

Instead of bootstrapping the data to estimate variance (EZ-MAP), theoretical adjustments can also be used to adjust z-scores across the reference and comparison groups to eliminate group-wise bias (DisCo-Z method). Potential benefits of this method include ease of implementation (single z-transformation), cost-effectiveness (does not require data collection on additional subjects), robustness (bias eliminated at less stringent alpha values of $P < 0.10$) and a more consistent/desired number of extremes regardless of sample size relative to independent samples approaches. The DisCo-Z adjustments for the reference and comparison groups also directly account for sample size separately, easily allowing for unmatched sample sizes that are common in neuroimaging studies. Similar to previous studies (Lipton et al., 2008; Patel et al., 2007), aspects of the DisCo-Z method can also be utilized for single subject data analyses.

The theoretical basis for the DisCo-Z assumes that the samples are derived from a normal distribution, an assumption which may be violated across different types of neuroimaging data. However, current results indicate that the proposed adjusted method was robust for eliminating bias in several distributions (chi-square and *t*-distributions) that deviated from normality both in terms of skew and kurtosis. As expected, the number of positive and negative extremes was still differentially affected by the degree of skew and kurtosis in our simulations, such that additional methods are needed to return a user-defined probability of extremes (i.e., 2% of total values) for these distributions.

Several other key observations were also apparent from simulations. First, the properties of the initial distribution (normal, kurtotic, or skewed) impacted both the resulting number of surviving extremes as well as the degree of bias. Specifically, the magnitude of bias was inversely related to the degree of skew (parametrically manipulated based on the df in the chi-squared distribution) in the direction of the skew. For example, the magnitude of bias was reduced for positive extremes in distributions with strong positive skew, whereas this trend was reversed in distributions with negative skew (Figs. 4 and 5). Similarly, a greater number of extreme voxels survived in the *t*-distribution relative to the normal distribution, which is expected due to kurtosis. Thus, it is critical that studies compare the average frequencies of SSA across patients and controls rather than relying on probability values alone based on a standard normal distribution. Second, the magnitude of bias was generally greater for the simulated data (i.e., more significant tests) versus real DTI data, which may have resulted from the additional constraint for spatially contiguous voxels (i.e., clusters) in the DTI data. Determination of cluster size for true false positive correction can be easily implemented with most standard neuroimaging packages based on Gaussian random field theory or Monte Carlo simulations. However, the characteristics of the SSA of interest must also be considered, as smaller abnormalities (e.g., petechial hemorrhages) may be difficult to detect using standard corrective thresholds dependent on image smoothness.

Results from several studies using z-transformation approaches that do not correct for potential bias in SSA have been previously reported in the neuropsychiatric literature, including two reports from our group. Specifically, we have previously reported increased FA during the semi-acute phase of mTBI using traditional ROI analyses in three independent samples of mTBI patients and HC (Ling et al., 2012; Mayer et al., 2010, 2012). However, the white matter tracts exhibiting evidence of increased FA were variable across our independent adult samples, leading us to adopt the pothole method as a potentially more sensitive method for detecting white matter injury (Ling et al., 2012). In contrast to our significant findings from pothole analyses, results from the DisCo-Z method were not indicative of significant group differences. However, consistent with our original results, the number of clusters with increased FA was decreased during the second visit, although the meaning of this finding is not clear given the non-significant Time 1 between-group results.

A reanalysis of the pediatric mTBI data using the DisCo-Z method indicated that our finding of an increased number of clusters of increased FA in the semi-acute injury phase remained statistically significant, and the longitudinal effects (no change) also remained the same. However, the magnitude of these group differences was smaller following appropriate correction with the DisCo-Z. Therefore, current and similar results recently obtained from another group (Watts et al., 2014) suggest that a careful reconsideration of the pothole method is warranted.

In summary, spatially heterogeneous white matter injuries in conjunction with strict corrections for reducing false positives may limit the utility of traditional analytic approaches for identifying the frequency of SSA in various neuropsychiatric populations. Although robust approaches for identifying SSA are needed, the statistical assumptions of these different approaches need to be carefully evaluated. For example, we have shown that the pothole method for SSA analyses introduces a systematic bias which over-estimates the number of extremes in the comparison group while underestimating extremes from the reference group. This bias can be corrected through the use of an independent reference group or the leave-one-out method using appropriately adjusted z-thresholds to account for differences in sample size. Alternatively, z-thresholds can be theoretically corrected to represent the true probability of extremes in each distribution or bootstrapped to estimate the variance (Kim et al., 2013). The DisCo-Z correction is cost effective, robust and relatively easy to implement using most statistical packages.

REFERENCES

Bazarian JJ, Zhu T, Blyth B, Borrino A, Zhong J (2012): Subject-specific changes in brain white matter on diffusion tensor imaging after sports-related concussion. Magn Reson Imaging 30:171–180.

Bouix S, Pasternak O, Rathi Y, Pelavin PE, Zafonte R, Shenton ME (2013): Increased gray matter diffusion anisotropy in patients with persistent post-concussive symptoms following mild traumatic brain injury. PLoS One 8:e66205.

Bulmer MG (1979): Principles of statistics. Mineola, NY: Dover Publications.

Cook RD, Weisberg S (1982): Residuals and influence in regression. NY: Chapman & Hall.

Cox R, Glen D (2006): Efficient, robust, nonlinear, and guaranteed positive definite diffusion tensor estimation. Seattle, WA: International Society for Magnetic Resonance in Medicine.

Cox RW (1996): AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Davenport ND, Lim KO, Armstrong MT, Sponheim SR (2012): Diffuse and spatially variable white matter disruptions are associated with blast-related mild traumatic brain injury. Neuroimage 59:2017–2024.

Davis KL, Stewart DG, Friedman JI, Buchsbaum M, Harvey PD, Hof PR, Buxbaum J, Haroutunian V (2003): White matter changes in schizophrenia: evidence for myelin-related dysfunction. Arch Gen Psychiatry 60:443–456.

Ding Z, Gore JC, Anderson AW (2005): Reduction of noise in diffusion tensor images using anisotropic smoothing. Magn Reson Med 53:485–490.

Ehrlich S, Geisler D, Yendiki A, Panneck P, Roessner V, Calhoun VD, Magnotta VA, Gollub RL, White T (2013): Associations of white matter integrity and cortical thickness in patients with schizophrenia and healthy controls. Schizophr Bull 40:665–674.

Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD (1996): Detecting activations in PET and fMRI: Levels of inference and power. Neuroimage 4:223–235.

Ge Y, Law M, Grossman RI (2005): Applications of diffusion tensor MR imaging in multiple sclerosis. Ann N Y Acad Sci 1064: 202–219.

Geisser S (1993): Predictive interference: An introduction. Boca Raton: CRC Press.

Hayasaka S, Phan KL, Liberzon I, Worsley KJ, Nichols TE (2004): Nonstationary cluster-size inference with random field and permutation methods. Neuroimage 22:676–687.

Hellyer PJ, Leech R, Ham TE, Bonnelle V, Sharp DJ (2012): Individual prediction of white matter injury following traumatic brain injury. Ann Neurol 73:489–499.

Hulkower MB, Poliak DB, Rosenbaum SB, Zimmerman ME, Lipton ML (2013): A decade of DTI in traumatic brain injury: 10 years and 100 articles later. AJNR Am J Neuroradiol 34: 2064–2074.

Jorge RE, Acion L, White T, Tordesillas-Gutierrez D, Pierson R, Crespo-Facorro B, Magnotta VA (2012): White matter abnormalities in veterans with mild traumatic brain injury. Am J Psychiatry 169:1284–1291.

Kim N, Branch CA, Kim M, Lipton ML (2013): Whole brain approaches for identification of microstructural abnormalities in individual patients: Comparison of techniques applied to mild traumatic brain injury. PLoS ONE 8:e59382.

Ling JM, Pena A, Yeo RA, Merideth FL, Klimaj S, Gasparovic C, Mayer AR (2012): Biomarkers of increased diffusion anisotropy in semi-acute mild traumatic brain injury: a longitudinal perspective. Brain 135:1281–1292.

Lipton ML, Gellella E, Lo C, Gold T, Ardekani BA, Shifteh K, Bello JA, Branch CA (2008): Multifocal white matter ultrastructural abnormalities in mild traumatic brain injury with cognitive disability: A voxel-wise analysis of diffusion tensor imaging. J Neurotrauma 25:1335–1342.

Lipton ML, Kim N, Park YK, Hulkower MB, Gardin TM, Shifteh K, Kim M, Zimmerman ME, Lipton RB, Branch CA (2012): 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 6:329–342.

Mac Donald CL, Johnson AM, Cooper D, Nelson EC, Werner NJ, Shimony JS, Snyder AZ, Raichle ME, Witherow JR, Fang R, Flaherty SF, Brody DL (2011): Detection of blast-related traumatic brain injury in U.S. military personnel. N Engl J Med 364:2091–2100.

Mayer AR, Ling J, Mannell MV, Gasparovic C, Phillips JP, Doezema D, Reichard R, Yeo RA (2010): A prospective diffusion tensor imaging study in mild traumatic brain injury. Neurology 74:643–650.

Mayer AR, Ling JM, Yang Z, Pena A, Yeo RA, Klimaj S (2012): Diffusion abnormalities in pediatric mild traumatic brain injury. J Neurosci 32:17961–17969.

Monnig MA, Tonigan JS, Yeo RA, Thoma RJ, McCrady BS (2013): White matter volume in alcohol use disorders: A meta-analysis. Addict Biol 18:581–592.

Mori S, van Zijl PC (2007): Human white matter atlas. Am J Psychiatry 164:1005.

Niogi SN, Mukherjee P (2010): Diffusion tensor imaging of mild traumatic brain injury. J Head Trauma Rehabil 25:241–255.

Pasternak O, Koerte IK, Bouix S, Fredman E, Sasaki T, Mayinger M, Helmer KG, Johnson AM, Holmes JD, Forwell LA, Skopelja EN, Shenton ME, Echlin PS (2014): Hockey Concussion Education Project, Part 2. Microstructural white matter alterations in acutely concussed ice hockey players: A longitudinal free-water MRI study. J Neurosurg 120:873–881.

Patel SA, Hum BA, Gonzalez CF, Schwartzman RJ, Faro SH, Mohamed FB (2007): Application of voxelwise analysis in the detection of regions of reduced fractional anisotropy in multiple sclerosis patients. J Magn Reson Imaging 26:552–556.

Poline JB, Mazoyer BM (1993): Analysis of individual positron emission tomography activation maps by detection of high signal-to-noise-ratio pixel clusters. J Cereb Blood Flow Metab 13:425–437.

Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW (2009): A new method for improving functional-to-structural MRI alignment using local Pearson correlation. Neuroimage 44:839–848.

Scheel M, Prokscha T, Bayerl M, Gallinat J, Montag C (2012): Myelination deficits in schizophrenia: Evidence from diffusion tensor imaging. Brain Struct Funct 218:151–156.

Shenton ME, Hamoda HM, Schneiderman JS, Bouix S, Pasternak O, Rathi Y, Vu MA, Purohit MP, Helmer K, Koerte I, Lin AP, Westin CF, Kikinis R, Kubicki M, Stern RA, Zafonte R (2012): A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. Brain Imaging Behav 6:137–192.

Viviani R, Beschoner P, Jaeckle T, Hipp P, Kassubek J, Schmitz B (2007): The bootstrap and cross-validation in neuroimaging applications: estimation of the distribution of extrema of random fields for single volume tests, with an application to ADC maps. Hum Brain Mapp 28:1075–1088.

Watts R, Thomas A, Filippi CG, Nickerson JP, Freeman K (2014): Potholes and Molehills: Bias in the diagnostic performance of diffusion-tensor imaging in concussion. Radiology 131856.

White T, Ehrlich S, Ho BC, Manoach DS, Caprihan A, Schulz SC, Andreasen NC, Gollub RL, Calhoun VD, Magnotta VA (2013): Spatial characteristics of white matter abnormalities in schizophrenia. Schizophr Bull 39:1077–1086.

White T, Schmidt M, Karatekin C (2009): White matter 'potholes' in early-onset schizophrenia: a new approach to evaluate white matter microstructure using diffusion tensor imaging. Psychiatry Res 174:110–115.