ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage xx (2006) xxx – xxx

Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters

SungWon Chung,[a,b] Ying Lu,[a,b] and Roland G. Henry[a,b,*]

[a]*Graduate Group in Bioengineering, University of California San Francisco and Berkeley, USA*
[b]*Center for Molecular and Functional Imaging, Department of Radiology, University of California San Francisco, San Francisco, CA, USA*

Received 2 May 2006; revised 20 June 2006; accepted 5 July 2006

Bootstrap is an empirical non-parametric statistical technique based on data resampling that has been used to quantify uncertainties of diffusion tensor MRI (DTI) parameters, useful in tractography and in assessing DTI methods. The current bootstrap method (repetition bootstrap) used for DTI analysis performs resampling within the data sharing common diffusion gradients, requiring multiple acquisitions for each diffusion gradient. Recently, wild bootstrap was proposed that can be applied without multiple acquisitions. In this paper, two new approaches are introduced called residual bootstrap and repetition bootknife. We show that repetition bootknife corrects for the large bias present in the repetition bootstrap method and, therefore, better estimates the standard errors. Like wild bootstrap, residual bootstrap is applicable to single acquisition scheme, and both are based on regression residuals (called model-based resampling). Residual bootstrap is based on the assumption that non-constant variance of measured diffusion-attenuated signals can be modeled, which is actually the assumption behind the widely used weighted least squares solution of diffusion tensor. The performances of these bootstrap approaches were compared in terms of bias, variance, and overall error of bootstrap-estimated standard error by Monte Carlo simulation. We demonstrate that residual bootstrap has smaller biases and overall errors, which enables estimation of uncertainties with higher accuracy. Understanding the properties of these bootstrap procedures will help us to choose the optimal approach for estimating uncertainties that can benefit hypothesis testing based on DTI parameters, probabilistic fiber tracking, and optimizing DTI methods.

Keywords: Diffusion tensor; MRI; Bootstrap; Monte Carlo simulation; Fiber tracking

Introduction

Diffusion tensor MRI (DTI) is a diffusion-weighted MRI technique capable of accurately describing anisotropic diffusion properties within a voxel (Basser et al., 1994a,b). DTI was a breakthrough in the studies of white matter microstructure through characterization with DTI parameters and delineation of white matter pathways with DTI fiber tracking. In order to conduct the statistical hypothesis tests on DTI parameters in different pathophysiologic conditions, especially for voxel-wise longitudinal studies, or to follow the white matter tracks in probabilistic sense, characterization of uncertainties associated with estimated DTI parameters is essential. One approach for this is an empirical, non-parametric statistical technique based on data resampling called bootstrap (Efron, 1979). Bootstrap was designed to replace complicated and often inaccurate approximations to uncertainty measures, such as biases and variances, by computer simulation based on real data. The bootstrap approach can be very helpful in DTI where final parameters of interest are known to be complicated nonlinear function of measured signals.

In DTI, a particular implementation of bootstrap was proposed (Pajevic and Basser, 2003) in which resampling was done within the data sharing common diffusion gradients. This approach makes no assumptions about the noise properties at the cost of requiring multiple acquisitions for each diffusion gradient; thus, we call this approach repetition bootstrap. Applications of repetition bootstrap were reported for the fiber tracking (Jones, 2003; Jones and Pierpaoli, 2005; Lazar and Alexander, 2005), quality assessment (Heim et al., 2004), and comparison of DTI anisotropy indices (Hasan et al., 2004). Unfortunately, a substantial underestimation bias in the degree of uncertainty was reported for this method, which degrades the reliability of bootstrap with small numbers of repeats (O'Gorman and Jones, 2005). Furthermore, a small number of samples is likely to be the case with the most applications, especially in clinical settings where acquisition time is limited.

In addition to the limitation of scan times, there is an interest in obtaining more diffusion sensitizing directions at the cost of repetitions. By definition, the repetition bootstrap approach cannot be used when only one measurement per each diffusion direction is

* Corresponding author. Center for Molecular and Functional Imaging, 185 Berry St., Suite 350, Box 0946, Department of Radiology, UCSF, San Francisco, CA 94107-0946, USA. Fax: +1 415 353 9425.
E-mail address: henry@mrsc.ucsf.edu (R.G. Henry).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.07.001

made. Acquisition of a single measurement is becoming more common practice with evidence that DTI parameters can be estimated more robustly with more diffusion gradient directions and with increasing interest in high angular resolution diffusion-weighted MRI (HARDI). In order to deal with the desire to acquire more diffusion sensitizing directions instead of multiple repetitions of the same directions, implementing wild bootstrap in the DTI analysis was proposed (Whitcher et al., 2005). Wild bootstrap is a model-based resampling technique designed to investigate the uncertainty in the linear regression with heteroscedasticity, i.e. non-constant variance with different regressors, of unknown form (Davison and Hinkley, 2003; Liu, 1988).

In this paper, we first describe the property of the downward bias of the estimated degree of uncertainty in the repetition bootstrap and propose to reduce this bias by implementing the bootknife approach (Hesterberg, 2004), which we call repetition bootknife. Evidence of bias correction actually improving the overall error of estimation is presented as well. Then, we investigate the feasibility of another model-based resampling approach called residual bootstrap, a well-known resampling technique in the statistics. Using Monte Carlo simulation, we compare the performance of repetition bootstrap, repetition bootknife, wild bootstrap, and residual bootstrap in terms of accuracy of estimating the degrees of uncertainty in diverse conditions such as different number of gradients, number of repetitions, diffusion tensor anisotropy, and partial volume with multiple tensors. Particular attention is paid to DTI sampling conditions within clinically feasible range since our ultimate goal is to establish the optimal bootstrap procedure that can be applied to clinical data. Based on the results, the optimal bootstrap approaches under various DTI sampling scheme are discussed.

Methods

Underestimation of standard errors by repetition bootstrap

Standard errors estimated by bootstrap are known to be generally smaller than the ideal values (downward biased) due to the basic mechanism of bootstrap. Bootstrap assumes that the empirical probability distribution $\hat{F}$, created by putting equal probabilities of $1/n$ to all the n elements of a sample, faithfully represents the unknown population probability distribution F from which the sample is drawn. Creating bootstrap samples from the sample $\hat{F}$ can be thought of as replicating the process of drawing new samples from the unknown population. Thus, an approximate distribution of some statistic $\hat{\theta}$ (some function of the sample $\hat{F}$ as an estimate of the parameter θ of the population F) can be generated via the bootstrap algorithm. If many samples of size n had been drawn from the population F, the standard deviation of the distribution of the statistic $\hat{\theta}$ would indicate the precision of $\hat{\theta}$; this is defined as the standard error of the statistic $\hat{\theta}$ of the sample $\hat{F}$. Thus, the standard error can be estimated by simply calculating the standard deviation of $\hat{\theta}^*$, which is the statistic of interest calculated from the bootstrap samples (Efron and Tibshirani, 1993).

When the original sample size n is small, bootstrap-estimated uncertainties are noticeably downward biased because the original sample that bootstrap relies on is biased. This phenomenon is similar to the bias in the estimator of population variance σ^2. It is well known that the estimator

$$s^2 = n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1}$$

is biased while the estimator

$$s^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{2}$$

is unbiased. This downward bias is a factor of $(n-1)/n$. When n is sufficiently large, this factor hardly makes any difference in the estimation, and the biased estimator is actually known to be slightly better than unbiased estimator in terms of mean squared error (MSE) defined by

$$\mathrm{MSE} = E((\hat{\theta}-\theta)^2) = \mathrm{bias}^2 + \mathrm{SD}^2 \tag{3}$$

$$\mathrm{bias} = E(\hat{\theta}) - \theta \tag{4}$$

$$\mathrm{SD} = \sqrt{(E(\hat{\theta}^2 - E^2(\hat{\theta}))} \tag{5}$$

where $\theta=\sigma$ and $\hat{\theta}=\hat{\sigma}$. Note that how far away estimator $\hat{\theta}$ is distributed from the population parameter θ depends on both the bias and the variability of the estimate (shown as SD or standard deviation in Eq. (5)) and the mean squared error (MSE) reflects the average quadratic loss or distance.

For the bootstrap, the usual estimator of uncertainty such as standard error can be thought to correspond to Eq. (1), the biased estimator. In particular, the bootstrap estimator of variance (squared standard error) of the sample mean is different from the unbiased estimator by the factor of $(n-1)/n$ (Efron and Tibshirani, 1993; Hesterberg, 2004). When bootstrap is performed on the samples from stratified random sampling, the bootstrap bias depends on the size of individual strata (corresponding to number of repetitions for repetition bootstrap), not the size of total sample (corresponding to number of repetitions times the number of gradient directions), which makes the bias substantial in situations with many small strata (Hesterberg, 2004). For instance, when the statistic being bootstrapped is a linear function of the means from multiple strata, the degree of bias for the bootstrap-estimated variance can be expressed as the scaling factor of $(n-1)/n$ just like the sample mean in the non-stratified sampling case described above, though now n is the number of samples in each stratum (Rao and Wu, 1988; Shao, 1996).

Repetition bootstrap can be regarded as an extreme case of stratified bootstrap in the sense that measurements with the same diffusion gradients (including $b=0$) are treated as strata and bootstrap resampling is performed only within each strata. Since it is unlikely that acquisitions will be repeated more than a few times even in experimental studies, repetition bootstrap will generally underestimate the standard error of DTI parameters to a substantial degree. DTI parameters are not linear functions of the raw measurements, and the degree of bias for bootstrap-estimated uncertainty is difficult to express analytically, though we might expect it to be somewhat around $\sqrt{(n-1)/n}$ for the standard error, where n is number of repetitions, not total number of measurements. Assuming that this is true, in repetitions of 2, 3, 4, and 5, we would expect the repetition bootstrap to estimate standard errors that are only about 71, 82, 87, and 89% of the true values.

Multiple algorithms have been proposed to correct this bias in the stratified sampling (Rao and Wu, 1988; Shao, 1996, 2003), and

Please cite this article as: SungWon Chung et al., Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.001.

in this paper, we propose a very simple modification of the conventional repetition bootstrap based on the bootknife algorithm (Hesterberg, 2004); thus we call this approach repetition bootknife. Bootknife is a resampling technique combining the features of jackknife and bootstrap as implied by name. Bootknife samples are created by first randomly omitting one sample from the original sample of size n in each stratum (jackknife) and drawing a bootstrap sample of size n with replacement from the remaining sample with size $n-1$ (bootstrap). The strata are the repeats for each diffusion gradient just like the repetition bootstrap originally proposed (Pajevic and Basser, 2003), and the rule that resampling does not mix the elements from different strata (gradients) is not violated just like repetition bootstrap (thus we will call these two algorithms collectively repetition-based or stratified resampling). Since bias correction might actually increase the MSE by increasing the variance of the standard error estimates more than the decrease in the bias, the total MSE needs to be compared with and without the bias correction. This will tell whether the bias correction is actually beneficial.

Residual bootstrap and wild bootstrap

Possible alternatives to repetition-based resampling are model-based resampling approaches such as the residual bootstrap and wild bootstrap. Implementation of the wild bootstrap was introduced (Whitcher et al., 2005) while there are no reports of residual bootstrap in DTI. Model-based resampling refers to the bootstrap resampling technique applied to the linear regression model, where the residuals based on the initially fitted model are resampled instead of the raw sample values. One might choose to do resampling pairs (of certain regressors and response) instead of residuals, but this approach, called pair bootstrap, would not be suitable in DTI since uncertainties estimated by pair bootstrap include variance generated due to different design (such as skipping some diffusion gradients) which does not reflect the fixed design of DTI. Thus, pair bootstrap is not considered in this study. Another possibility is to assume symmetry in the distribution of residuals for a given data point, and resample based on randomly changing the signs of the residuals; this is an implementation of the wild bootstrap. A third alternative is to assume that all residuals have similar distributions and freely resample among them without stratification; this is called residual bootstrap.

Since model-based resampling is 'based on a model', the model (diffusion tensor) needs to be adequate in describing the measured diffusion signals so that the error terms at different design points (different gradient directions for example) will have a common mean of zero. Regarding the variance of errors, residual bootstrap can be used in the homoscedastic condition (constant variance of error terms for the different design points) and also in the heteroscedastic condition as long as the heteroscedasticity can be modeled. If heteroscedasticity cannot be described mathematically, wild bootstrap may be a better approach since it does not require homoscedasticity.

In DTI, it was recognized from the beginning that the degree of uncertainty of log transformed signals used for linear regression is the inverse of the raw signals, and this property has been widely used for constructing weighting factors in the weighted least squares solution of diffusion tensor (Basser et al., 1994a). Similarly, the residual bootstrap of DTI can be carried out based on the propagation of variance in log transformed signals, and the details of residual bootstrap as well as the diffusion tensor calculation are described in the following.

In a DTI experiment, the diffusion-weighted signal S is modeled by

$$S(\mathbf{g}_j) = S_0 \exp(-b\mathbf{g}_j^T \mathbf{D}\mathbf{g}_j), \text{ with } j = 1, 2 \ldots N, \tag{6}$$

where S_0 is the signal intensity without diffusion weighting, b is the diffusion weighting factor, $\mathbf{D}$ is effective self-diffusion tensor in the form of 3×3 positive definite matrix, $\mathbf{g}$ is 3×1 unit vector of the diffusion-sensitive gradient direction, and N is the total number of experiments, including repeated measurements. By log transform, the equation above becomes

$$\ln(S(\mathbf{g}_j)) = \ln(S_0) - b\mathbf{g}_j^T \mathbf{D}\mathbf{g}_j, \tag{7}$$

which can be structured into well-known multiple linear regression form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{8}$$

where $\mathbf{y} = [\ln(S(\mathbf{g}_1)), \ln(S(\mathbf{g}_2)), \ldots, \ln(S(\mathbf{g}_N))]^T$ are the logarithm of measured signals, $\boldsymbol{\beta} = [D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{xz}, D_{yz}, \ln S_0]^T$ are the unknown regression coefficients including the 6 unique elements of $\mathbf{D}$, $\mathbf{X}$ is a design matrix of different diffusion gradient directions,

$$\mathbf{X} = -b \begin{bmatrix} g_{1x}^2 & g_{1y}^2 & g_{1z}^2 & 2g_{1x}g_{1y} & 2g_{1x}g_{1z} & 2g_{1y}g_{1z} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{Nx}^2 & g_{Ny}^2 & g_{Nz}^2 & 2g_{Nx}g_{Ny} & 2g_{Nx}g_{Nz} & 2g_{Ny}g_{Nz} & 1 \end{bmatrix},$$

and $\boldsymbol{\varepsilon} = [\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_N]^T$ are error terms. The weighted least squares (WLS) estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y}. \tag{9}$$

In order to determine the diagonal weighting matrix $\mathbf{W}$, the ordinary least squares (OLS) estimate is calculated first by $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ leading to OLS fitted log measurements $\hat{\boldsymbol{\mu}}_{\mathrm{OLS}} = \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ and fitted diffusion signals $\hat{\mathbf{S}}_{\mathbf{g}} = \exp(\hat{\boldsymbol{\mu}}_{\mathrm{OLS}})$. Then, $\mathbf{W} = \mathrm{diag}(\hat{\mathbf{S}}_{\mathbf{g}}^2)$, which is based on the property

$$\mathrm{Var}(\varepsilon_j) = \sigma^2 / S_j^2 \tag{10}$$

where σ is the standard deviation of noise in the raw signal. σ is assumed to be constant for each voxel regardless of the measured signal intensity.

After $\hat{\boldsymbol{\beta}}$ is calculated, the WLS fitted log measurements $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ are used to calculate the residual vector $\mathbf{e} = \mathbf{y} - \hat{\boldsymbol{\mu}}$. In order to resample the errors, error terms ε_j need to be i.i.d. (independent and identically distributed) to satisfy the basic assumption of bootstrap that the samples are i.i.d. However, generally, the raw residuals $\mathbf{e}$ do not satisfy this condition due to the effect of possible heterogeneous leverages for different points. Furthermore, ε_j actually have non-constant variance (heteroscedasticity) but for DTI this can be modeled as shown in Eq. (10). Therefore, raw residuals $y_j - \hat{\mu}_j$ need to be modified to have constant variance by following equation

$$r_j = \frac{y_j - \hat{\mu}_j}{w_j^{-1/2}(1 - h_j)^{1/2}}, \tag{11}$$

Please cite this article as: SungWon Chung et al., Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.001.

where the weighting factor w_j is the jth diagonal element of $\mathbf{W}$ and the leverage value h_j is the jth diagonal element of the hat matrix $\mathbf{H}$ defined by $\mathbf{H}=\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}$. Finally, residual bootstrap resampling is defined as

$$y_j^* = \mathbf{x}_j\hat{\boldsymbol{\beta}} + w_j^{-1/2}\varepsilon_j^*, \tag{12}$$

where y_j^* is jth element of resampled log measurements, $\mathbf{x}_j$ is the jth row of $\mathbf{X}$, and ε_j^* is randomly resampled with replacement from the set of centered modified residuals $r_1-\bar{r}$, $r_2-\bar{r}$, …,$r_N-\bar{r}$ (Davison and Hinkley, 2003).

A bootstrap sample set $\mathbf{y}^*=[y_1^*, y_2^*, \ldots, y_N^*]^T$ undergoes the WLS fitting procedure described above which leads to $\mathbf{D}^*$, from which a DTI parameter $\hat{\theta}^*$ such as FA (fractional anisotropy) is calculated. Resampling $\varepsilon^*=[\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_N^*]^T$ and calculating $\hat{\theta}^*$ are repeated for some fixed large number N_B (typically hundreds to thousands times) to acquire N_B independent bootstrap samples $\hat{\theta}^{*b}$, $b=1, 2, \ldots, N_B$. Here, the sample statistic $\hat{\theta}$ is an estimation of the true unknown θ (such as the noise-free FA of the voxel) using the original sample $\mathbf{y}$ by WLS, and $\hat{\theta}^*$are bootstrap replications of $\hat{\theta}$. The bootstrap-estimated standard error of $\hat{\theta}$ is simply the standard deviation of the N_B replications

$$\hat{se}_B = \left\{\sum_{b=1}^{B}[\hat{\theta}^*(b)-\hat{\theta}^*(\cdot)]^2/(N_B-1)\right\}^{1/2}, \tag{13}$$

where $\hat{\theta}^*(\cdot)=\Sigma_{b=1}^{B}\hat{\theta}^*(b)/N_B$.

As mentioned above, wild bootstrap is suitable when heteroscedasticity cannot be modeled. In DTI with least squares estimation, this means that we are not relying on Eq. (10) to modify raw residuals and resample residuals gathered from total design. Instead of resampling residuals from the pool of modified residuals causing the residuals from diffusion weighting of specific direction to be randomly distributed on any other directions, wild bootstrap creates variability by simply multiplying the individual residuals with a mutually independent random function. Wild bootstrap resampling is defined as

$$y_j^* = \mathbf{x}_j\hat{\boldsymbol{\beta}} + \varepsilon_j^* \tag{14}$$

where the resampled error ε_j^* is

$$\varepsilon_j^* = \frac{y_j-\hat{\mu}_j}{(1-h_j)^{1/2}}t_j \tag{15}$$

and t_j is i.i.d. random variables with $E(t_j)=0$, and $E(t_j^2)=1$. Commonly t_j is a two-point distribution, and in this study, the Rademacher distribution F2 with the property of $\Pr(t_j=1)=0.5$ and $\Pr(t_j=-1)=0.5$ was used due to its good performance (Davidson and Flachaire, 2001). Simply speaking, modified residuals are randomly multiplied by either +1 or −1 and then added back to the fitted point where they originated from, without being distributed to other design points. All the other steps are equivalent to residual bootstrap.

Monte Carlo simulation

The performances of the four bootstrap approaches in terms of bias, standard deviation, and overall error (MSE) of bootstrap-estimated standard error (or 95% confidence interval for the angle of primary eigenvector) were compared under diverse conditions by Monte Carlo simulation. We assumed that Rician noise is the only source of uncertainty in the diffusion signals. Schemes with the number of diffusion encoding directions, ranging from 6 to 54 were investigated. The six directions cases were based on the dual gradient scheme while the other number of directions was based on the electrostatic repulsion scheme (Jones, 2004). Two b value experiments were used, $b=0$ s/mm^2 and $b=1000$ s/mm^2 (or 3000 s/mm^2 when specified). Number of images for $b=0$ and $b>0$ was kept in the ratio of 1:6, such that for 54 directions, there were 9 $b=0$ images. One to 9 numbers of repetitions were studied since clinical DTI scans are rarely repeated 10 or more even with only 6 directions.

Simulation was performed in a similar manner as described elsewhere (Pierpaoli and Basser, 1996), using custom software in IDL 6.1 (Research Systems, Inc., Boulder, CO). After an ideal, noise-free diffusion tensor was derived based on the desired DTI parameters such as FA (0.2, 0.5, and 0.8 were considered) and D_{av} (=Tr(**D**)/3, fixed to 0.7×10^{-3} mm^2/s), noise-free diffusion-weighted signals along specific direction of diffusion gradients were calculated according to the Eq. (6) (S_0 arbitrarily set to 100). Then, noise modeled as complex random number with real and imaginary parts following Gaussian distribution of zero mean and standard deviation σ ($=S_0$/SNR) was added to the noise-free signal and the magnitude of the noisy signal was calculated. SNR of each $b=0$ image was set to 25 for this study unless specified otherwise. After a complete set of noisy signals was acquired, noisy diffusion tensor and DTI parameters were calculated. These steps were repeated a large number of times (100,000 used in this study), and a gold standard version of the standard error (or confidence interval of angle of primary eigenvector) for the DTI parameter of interest was directly calculated from the standard deviation (or 95% range of angle of primary eigenvector) of all the noisy parameters.

We also investigated the conditions with partial volume effects (PVE) where signals are actually originating from a system more complicated than a single tensor. Multiple regions in the brain are known to have PVE due to intravoxel crossing of two distinct axonal bundles, and this can violate the assumption of appropriateness of the single tensor DTI model in model-based resampling. Model-based resampling may not perform optimally with PVE, and therefore the performance of residual or wild bootstrap with PVE can be important when implementing bootstrap in clinical data. In this study, we focused on a mixture of white matter bundles (Alexander et al., 2001) where the diffusion-weighted signals come from two compartments described as

$$S(\mathbf{g}_j) = S_0 f\exp(-b\mathbf{g}_j^T\mathbf{D}_1\mathbf{g}_j) + S_0(1-f)\exp(-b\mathbf{g}_j^T\mathbf{D}_2\mathbf{g}_j) \tag{16}$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ represent the tensor from each compartment, f and $(1-f)$ are the signal fractions from $\mathbf{D}_1$ and $\mathbf{D}_2$. We assumed no exchange between the compartments, which will make PVE most pronounced. $\mathbf{D}_1$ and $\mathbf{D}_2$ were assumed to be prolate tensors with FA=0.7, f was fixed to 0.5, and angles between primary eigenvectors of two tensors were varied. Then, the usual single tensor design matrix was used to fit the diffusion signals, in the calculation of gold standard or bootstrap estimates of SE. We also created the equivalent single tensor system as follows. The noise-free PVE data were fitted, and the calculated tensor was used to define the equivalent single tensor. Noise was then added to this equivalent tensor for further analysis.

In order to evaluate the performance of the bootstrap approaches, the bootstrap procedures described above were

Please cite this article as: SungWon Chung et al., Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.001.

performed either directly on the diffusion-weighted signals (repetition bootstrap and repetition bootknife) or on the residuals between measured signals and fitted signals (residual bootstrap and wild bootstrap) to create the bootstrap samples, and they were used to calculate the diffusion tensors and the DTI parameters of interest. This process was repeated N_B times (1000 was used for this study), and finally the bootstrap-estimated standard errors were calculated by Eq. (13). Since standard errors cannot be used for vector quantities such as the primary eigenvector, the 95 percentile confidence intervals of the minimum angle between each bootstrap estimate and the average primary eigenvector were used instead (Jones, 2003). Though not used in this study, there is an alternative measure of accuracy of the primary eigenvector based on the dispersion parameter of Watson distribution (Schwartzman et al., 2005). The experiment of bootstrap estimation of uncertainty was repeated 1000 times from which the variance of these estimations was calculated by Eq. (5), and the biases determined by comparing the expectation of estimations with the gold standard value by Eq. (4). The MSE of the bootstrap standard error estimates was computed from the biases and variances as shown in Eq. (3), which reflects the overall degrees of error and is an objective index of performance.

Results

Bias of repetition bootstrap

The typical downward bias of repetition bootstrap for small numbers of repetitions is demonstrated in Fig. 1a. Simulation results were acquired for varying number of repetitions, while the number

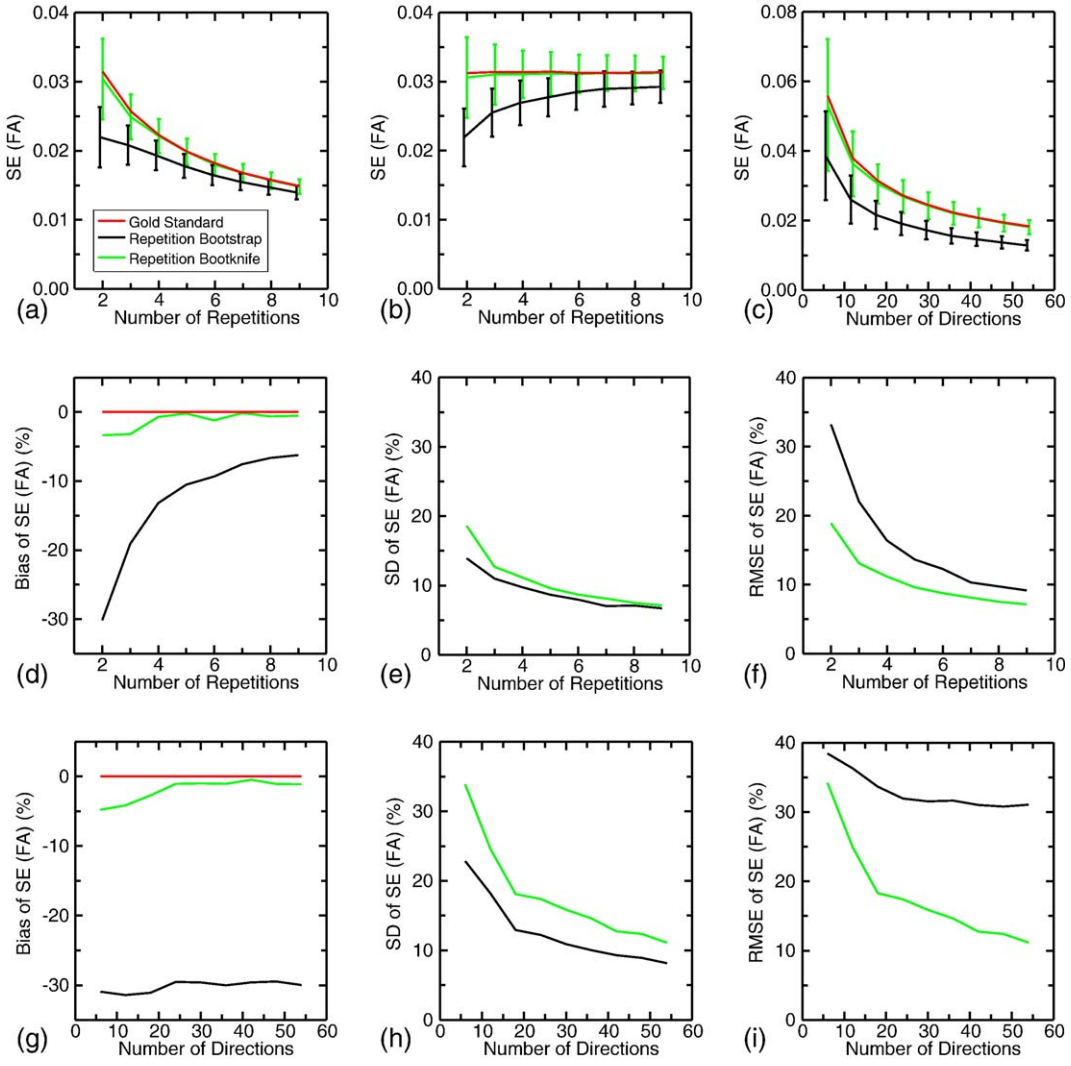

Fig. 1. (a) Standard error (SE) of FA estimated by repetition bootstrap (black line) and repetition bootknife (green line) plotted as mean and standard deviation (SD) in vertical bars with varying number of repetitions while the number of directions is fixed to 18. The gold standard SE is shown in red. Noisy diffusion signals were created by adding noise corresponding to SNR = 25 to noise-free signals from a prolate tensor with FA = 0.5, $D_{av} = 0.7 \times 10^{-3}$ mm^2/s. (b) Same as (a) except that SNR for individual acquisition is adjusted to keep the effective SNR after combining repetitions to be constant. SNRs used per acquisition in the repetitions from 2 to 9 are approximately 25.0, 20.4, 17.7, 15.8, 14.4, 13.3, 12.5, and 11.8. (c) Same as (a) except that number of directions is varied while the number of repetition is fixed to 2. Noise was added in the same way as in (a). (d–f) Bias, SD, and square root of MSE (RMSE) of bootstrap estimates of SE in % of gold standard SE for the data displayed in (a). % Bias, SD, and RMSE for the data displayed in (b) are almost identical to (d–f). (g–i) Bias, SD, and RMSE of bootstrap estimates of SE in % of gold standard SE for the data displayed in (c).

Please cite this article as: SungWon Chung et al., Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.001.

of gradient directions was fixed to 18 plus 3 $b=0$ images. Fig. 1a shows the mean and standard deviation of the bootstrap estimates of SE of FA (from 1000 experiments) as well as the gold standard value of SE (from 100,000 experiments). The mean of the repetition bootstrap SE estimates was substantially smaller than the gold standard, while the repetition bootknife estimates are nearly unbiased, though with slightly larger standard deviations than the repetition bootstrap. Figs. 1d–f show the bias, standard deviation (SD), and square root of mean squared error (RMSE) of the bootstrap estimates separately, displayed as a percent of the gold standard SE of FA value. This normalization allows the bootstrap performance to be compared to other conditions such as different DTI parameters, number of repetitions, and number of directions. Note, again, that repetition bootstrap was substantially downward biased down to 30% while the repetition bootknife was nearly unbiased. The repetition bootknife proved to be a better estimator with the smaller RMSEs, especially for small numbers of repetitions.

Figs. 1a and d–f also show that both repetition methods are more accurate with more repetitions. This simply reflects the fact that bootstrap performs better with a larger sample pool and in particular the estimates do not improve due to the increasing total SNR associated with more repeated acquisitions. To illustrate this point, the total SNR was held fixed for the different numbers of repetitions and is shown in Fig. 1b. For Fig. 1a, on the other hand, SNR of each repetition is fixed to 25 resulting in increasing SNR with more repetitions (and subsequent reduction in the gold standard SE). Fig. 1b shows that now the gold standard SE (FA) is constant instead of decreasing but the bootstrap bias, SD, and RMSE (% of gold standard) are almost identical to Figs. 1d–f (thus result not shown in the format of Figs. 1d–f), illustrating that the estimates are still more accurate with larger repetitions independent of total SNR. This clearly shows that SNR itself is not a factor influencing the bias of the repetition methods. For repetition bootstrap, the RMSE is primarily influenced by the decrease in bias with increasing number of repetitions. Given that the origin of the bias is due to small sample sizes, it is clear that increasing the number of samples (and not the SNR) determines the percent RMSE. For the repetition bootknife, the reduction in SD of the SE estimates with increasing repetitions is the strongest factor in the RMSE. In both repetition methods, the decreases in the SD of SE with increasing the number of repetitions are due to the increased sample size.

As expected, the underestimation bias of repetition bootstrap is problematic when the number of directions rather than the number of repetitions is increased. Note from Figs. 1c and g–i that, as more directions are acquired while the number of repetitions is fixed to 2, the degree of bias for repetition bootstrap hardly improves, which is expected since repetition bootstrap bias depends only on the number of repetitions. This property leads to a poor improvement of RMSE for repetition bootstrap even with a large number of directions. The repetition bootknife, on the other hand, has a small bias that becomes even smaller with more directions, leading to the similar trend of RMSE as Fig. 1f. Thus, the gap of performance between these two methods is more pronounced with a larger sample pool made from increasing the number of directions.

Bootstrap methods in the diffusion signals from single tensor model

Fig. 2 shows the performance of the four bootstrap approaches for estimating the 95% confidence interval (CI) of the angle of primary eigenvector with the number of repetitions between 1 and 6 and the number of directions fixed to 18. Since model-based resampling methods such as residual and wild bootstrap do not depend on repeated acquisitions, bootstrap performance can be

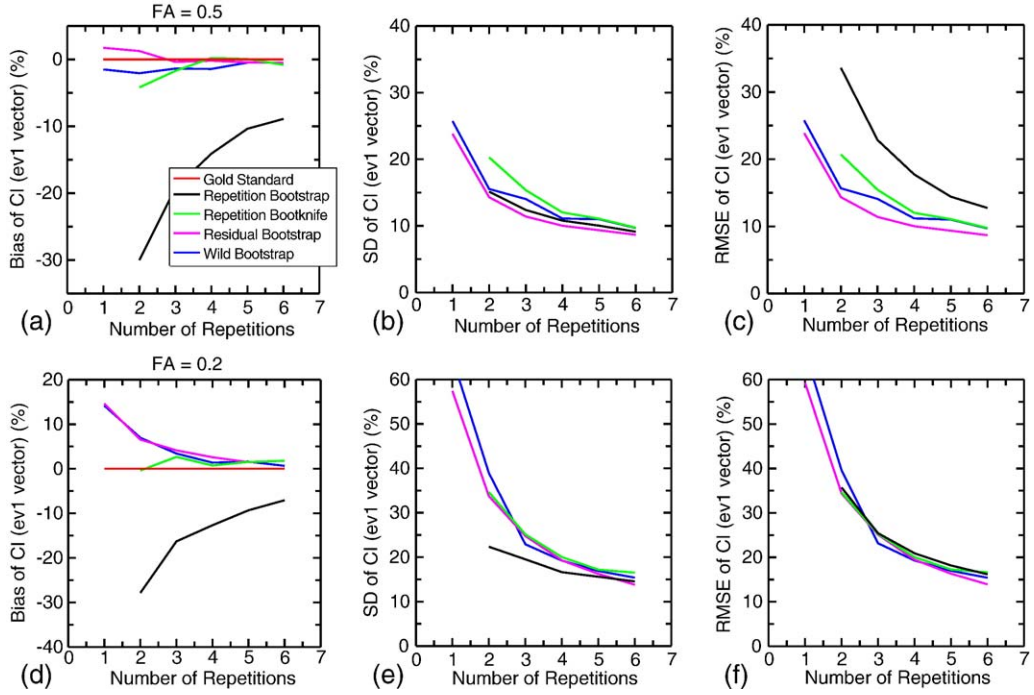

Fig. 2. Bias, SD, and RMSE of the 95th percentile confidence interval of the angle of primary eigenvector estimated by bootstrap methods with different number of repetitions while number of directions is fixed to 18. The DTI models were prolate tensors with FA of 0.5 (a–c) and 0.2 (d–f) and D_{av} of 0.7×10^{-3} mm^2/s for all. Data with FA of 0.8 have almost identical plots to (a–c).

Please cite this article as: SungWon Chung et al., Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.001.

shown even in the case of only one acquisition unlike the repetition methods for which results are displayed with number of repetitions starting at two. Figs. 2a–c show the bias, SD, and RMSE when the noise-free modeled diffusion tensor has a moderate anisotropy of FA=0.5. The repetition methods have a very similar pattern of bias, SD, and RMSE to that found for FA in Figs. 1d–f. Repetition bootstrap is substantially downward biased, and the overall error is smaller with repetition bootknife. The residual bootstrap and wild bootstrap methods are shown to be nearly unbiased, have small SD, and have RMSE smaller than the repetition methods. The residual bootstrap has slightly lower bias and SD than the wild bootstrap. For each value of the number of repetitions, residual bootstrap has the smallest RMSE, followed by the wild bootstrap, then repetition bootknife, and lastly repetition bootstrap. This bootstrap performance pattern is nearly identical in high anisotropy of FA=0.8 (result not shown), while in low anisotropy of FA=0.2 (Figs. 2d–f), all bootstrap methods suffer from worse performance. Figs. 2d–f show that the residual and wild bootstrap methods overestimate the CI especially for small numbers of repetitions, and the estimates of all the bootstrap methods are more dispersed at low FA than the estimates in medium or high anisotropy (note the scale difference of y axis). This implies that not only the primary eigenvector direction is more uncertain in the low anisotropy but also the ability of bootstrap to estimate the increased uncertainty is worse. Interestingly, all four bootstrap methods show similar RMSE in low anisotropy.

The bootstrap performance for estimating the primary eigenvector angle CI when the number of diffusion gradient directions is increased while the number of acquisitions is fixed to two is demonstrated in Fig. 3. As pointed out in Figs. 1c and g–i, repetition bootstrap bias remains relatively independent of the number of

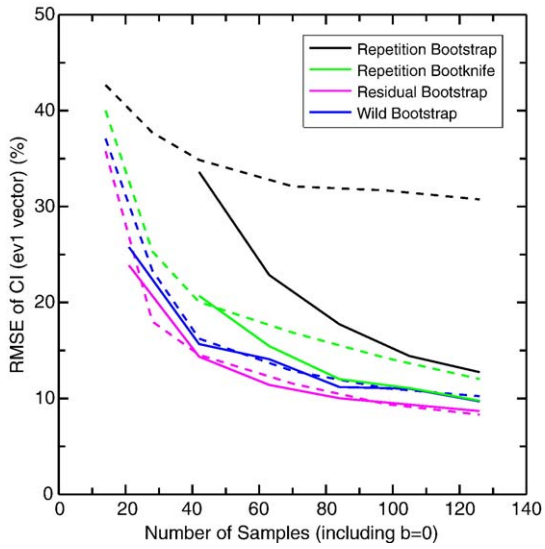

Fig. 4. Comparison of the RMSE of the 95th percentile confidence interval of the angle of the primary eigenvector estimated by bootstrap methods for varying numbers of repetitions and directions. Solid lines are results with different number of repetitions (ranging from 1 to 6) while number of directions is fixed to 18. Dashed lines are results with different number of directions (ranging from 6 to 54) while number of repetitions is fixed to 2. Noisy diffusion signals are created by adding noise corresponding to SNR=25 to noise-free signals from a prolate tensor with FA=0.5, $D_{av}=0.7\times10^{-3}$ mm^2/s. That is, solid line results are same as Fig. 2c while dashed line results are same as Fig. 3c. The number of samples includes $b=0$ measurements (with the number 1/6 of different number of diffusion directions).

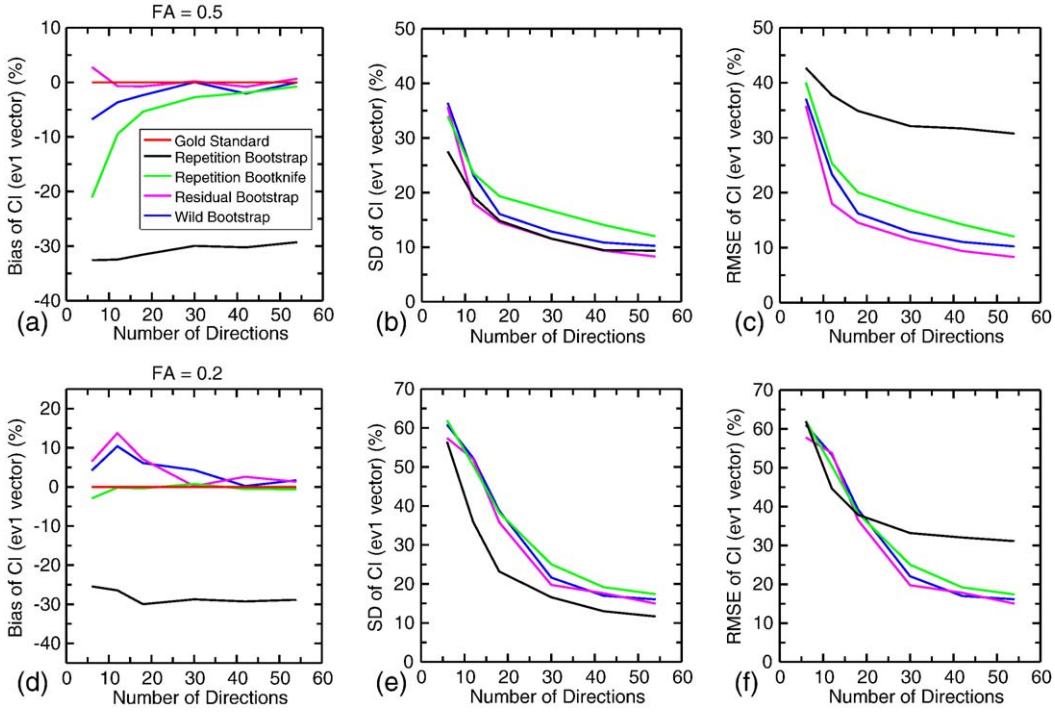

Fig. 3. Bias, SD, and RMSE of the 95th percentile confidence interval of the angle of the primary eigenvector estimated by bootstrap methods with different numbers of directions while the number of repetitions is fixed to 2. The DTI models were prolate tensors with FA of 0.5 (a–c) and 0.2 (d–f) and D_{av} of 0.7×10^{-3} mm^2/s for all. Data with FA of 0.8 have almost identical plots to (a–c).

Please cite this article as: SungWon Chung et al., Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.001.

directions, leading to substantially larger RMSE for large numbers of directions. The other three methods show similar trends of bias, SD, and RMSE to Fig. 2. The residual bootstrap is generally the least biased and variable followed by the wild bootstrap and repetition bootknife. Thus, residual bootstrap seems to have better overall performance than the others. Just as in Fig. 2, results for high anisotropy of FA=0.8 are almost identical to moderate anisotropy of FA=0.5, and for low anisotropy of FA=0.2, the model-based resampling shows some overestimation at lower numbers of directions that rapidly disappears with more directions. All bootstrap methods except for the repetition bootstrap show relatively small differences in RMSE in low anisotropy as well. When the data are acquired only once (meaning that number of repetition is one), repetition bootstrap and repetition bootknife are no longer available, while model-based resampling can still be used. Results for the performance of residual and wild bootstrap without any repeated acquisition (not shown) indicate that the trend is very similar to Fig. 3 though the bias, SD, and RMSE are larger than that of Fig. 3 since the sample size is only half of Fig. 3. Overall, residual bootstrap is less biased, less variable, and has smaller RMSE.

Fig. 4 shows Figs. 2c and 3c plotted together with the common x axis representing the number of samples (includes $b=0$) in order to clearly demonstrate the increase of sample size either by number of repetitions (solid lines) or directions (dashed lines). This shows that the residual bootstrap and wild bootstrap have a very similar trend of improvement of RMSE when either number of repetitions or directions is increased. As expected, the repetition bootstrap, on the other hand, does not benefit from increasing number of directions as much as number of repetitions. Even the repetition bootknife has a slight tendency of better performance with larger repetitions rather than directions.

Bootstrap methods in the diffusion signals from tensor mixture model

Fig. 5 shows the performance of the bootstrap methods when acquisitions are repeated one or two times with a relatively large number of diffusion directions of 54. This sampling scheme was chosen since the model (diffusion tensor) insufficiency is more likely to be an issue with the large number of directions, and clinical scans are not likely to be repeated more than one or two times with large number of directions. Typical diffusion weighting of $b=1000$ s/mm^2 was used, thus representing the scenario where resolving the PVE such as intravoxel crossing is not necessarily of primary interest. In order to separate the influence of PVE versus simply different diffusion tensor shape on the bootstrap, two separate results from different modeling are simultaneously displayed. Solid lines are results from modeling tensor mixture, while dashed lines are results from modeling single tensor equivalent to the fitted tensor to noise-free signals from tensor mixture. Of course, once the noisy diffusion signals are acquired in either way, then fitting a single tensor to the data is assumed, just as almost all the diffusion tensor analysis of real data is done (i.e. without applying HARDI or multiple tensor modeling of diffusion signals). In the case of two repetitions shown in Figs. 5a–c, all bootstrap methods show relatively small differences between the solid and dashed lines implying that PVE is not a significant factor in the performance of any of bootstrap methods. This trend is replicated in the case with only one repetition in Figs. 5d–f, where

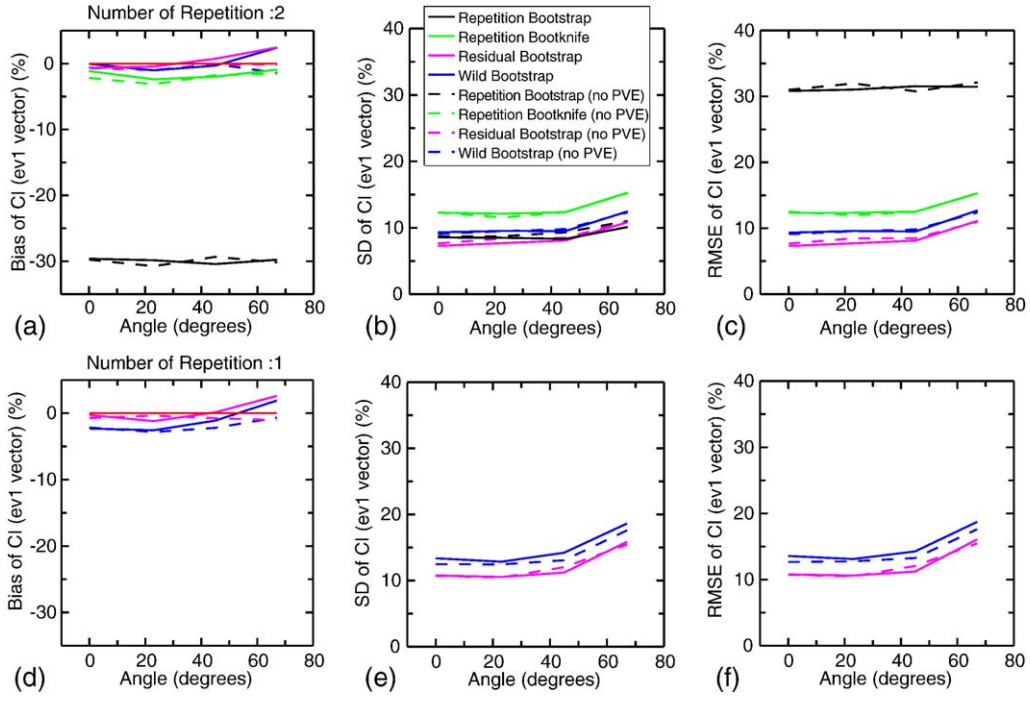

Fig. 5. Bias, SD, and RMSE of the 95th percentile confidence interval of the angle of the primary eigenvector estimated by bootstrap methods with varying angles between two tensors within a voxel for two (a–c) and one (d–f) repetitions while number of directions is fixed to 54. The primary eigenvectors of the two prolate tensors, each with FA=0.7 and $D_{av}=0.7\times10^{-3}$ mm^2/s, were positioned at angles of 0, 23, 45, and 68°. The solid lines are the usual single tensor fits to these modeled PVE. The dashed lines are single tensor fits to an equivalent single tensor (with the same FA, D_{av}, single tensor shape, etc.) found from the single tensor fit to the noise-free partial volume model and then refitted with noise added.

Please cite this article as: SungWon Chung et al., Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.001.

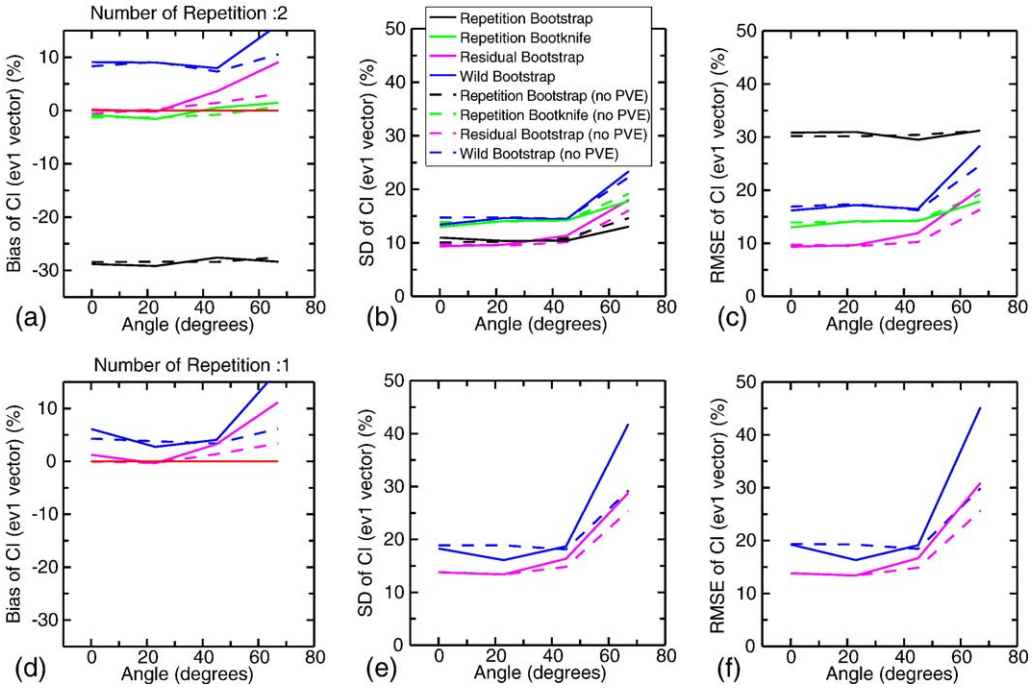

Fig. 6. Same as Fig. 5 except that *b* value used is 3000 s/mm^2.

the model-based resampling shows similar trends between solid and dashed lines.

Fig. 6 shows the results when the *b* value is increased to 3000 s/mm^2 for more pronounced PVE (Alexander et al., 2001). When the angles between primary eigenvectors of the two modeled tensors are small, the difference between solid and dashed lines is minimal, but when the angle is large such as 68°, solid lines of residual and wild bootstrap show larger bias and SD leading to larger RMSE. This supports our theory that model-base resampling is more susceptible to PVE. On the other hand, the repetition bootstrap and repetition bootknife methods are insensitive to the presence of PVE. Even so, the residual bootstrap still has smaller RMSE than the other methods even with the presence of PVE, except for the tensor mixture with 68° where the repetition bootknife seems to be slightly better than residual bootstrap.

Discussion

Bootstrap is a powerful method of estimating the uncertainties in DTI derived parameters, and it has been successfully implemented and shown to be useful in diverse applications such as probabilistic fiber tracking and quality assessment of DTI acquisitions. It also has the potential to be used for statistical tests such as voxel-based (or ROI-based) analysis of longitudinal (acquired at multiple time points) and cross-sectional DTI data. So far, only one particular implementation of bootstrap (repetition bootstrap) has been used in applications, but it is important to point out that bootstrap is not defined in a unique way, but rather a group of diverse algorithms sharing the basic concept. In this paper, we implemented four DTI bootstrap approaches, including two previously unreported, and tested them by Monte Carlo simulation under diverse conditions in search of the optimal method that can calculate the uncertainty reliably.

We showed that repetition bootstrap is substantially downward biased and introduced the repetition bootknife that successfully reduced the bias and mean squared error. We also introduced the residual bootstrap as another model-based resampling technique and compared the four bootstrap methods (repetition bootstrap, repetition bootknife, residual bootstrap, and wild bootstrap) for their performance in terms of bias, variance, and mean squared error. Our simulations demonstrated that, in the cases where DTI was acquired multiple times permitting all four bootstrap methods, model-based resampling outperforms repetition-based resampling if the model is true, suggesting that, even if multiple acquisitions exist, model-based resampling might be the better choice. When data are acquired only once with possibly many different diffusion encoding directions, repetition-based bootstrap is not feasible but model-based resampling can still be used. This allows greater freedom for researchers and/or clinicians in choosing a diffusion gradient sampling scheme when they are considering implementing bootstrap for DTI data. Within model-based resampling, residual bootstrap was consistently better than wild bootstrap, especially when the single tensor model was not sufficient. For repetition-based resampling, the modified version introduced here proved to be better than the original version.

Another important result of this study is that, with model-based bootstrap techniques, one can benefit from increasing number of directions just as much as increasing number of repetitions. Pajevic and Basser (2003) postulated that with many distinct non-collinear directions fewer repetitions would be required to achieve the same reliability of bootstrap, but their data indicated that the relationship between the number of directions, repetitions, and the quality of bootstrap estimates was somewhat complicated. For instance, their data show that CV of SE (RA) (similar to SD of SE (FA) (%) in our data) can actually increase with very large number of directions with fixed number of repetitions and that CV of SE (Trace) actually gradually worsens with more directions with fixed repetitions. We

Please cite this article as: SungWon Chung et al., Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.001.

believe that this is related to fixing number of $b=0$ images to one instead of increasing it to keep the ratio of $b=0$ to $b>0$ constant (such as 1:6 in this study). When we fixed the number of $b=0$ images to one regardless of the number of directions, we observed very similar trend that increasing the number of directions does not consistently improve the bootstrap performance (result not shown here). This effect is probably due to the strong leverage that the $b=0$ data have on the least squares fit when only one $b=0$ data is acquired. The influence of a data point on the fit depends on the leverage and variance of the data point compared to the others. When many encoding directions are used, the influence for a single $b=0$ data point can be large due to the large leverage. Our data with fixed ratio of $b=0$ to $b>0$ indicate that model-based resampling with either large number of directions or large number of repetitions has very similar performance, which was clearly demonstrated in Fig. 4. As long as model-based resampling is used and $b=0$ images are increased accordingly with more diffusion directions, total sample pool size alone determines the bootstrap performance. For repetition bootstrap, using more repetitions is always better than increasing the number of directions (even with increased number of $b=0$ images) since increasing the number of directions does not directly increase the resample pool size.

It is important to emphasize that, for bootstrap to be reliable, the sample pool size should be large enough, though with model-based techniques repetition is not a requirement anymore. Residual or wild bootstrap can generate estimates of SE of DTI parameters in a single acquisition, but unless the number of directions is large, bootstrap estimates will be highly variable. It is difficult to generalize how large the total sample pool should be because bootstrap performance depends on tensor anisotropy and shape and the DTI parameter of interest, and the definition of good performance depends on the sensitivity needed in the application. However, given an effect size, the bootstrap data will enable power calculations. This study shows how different bootstrap methods can perform under a few selected demonstrative conditions, but more studies are needed in order to have a more complete picture of bootstrap performance issue.

The results shown in this study were focused on the performance of bootstrap in estimating the uncertainty of primary eigenvector and FA because incorporating bootstrap to fiber tracking is of a great interest, but similar results were obtained for the eigenvalues and D_{av}. However, it is not clear how the bootstrap methods perform in estimating the entire probability density function (pdf) of DTI parameters except for some evidence of repetition bootstrap properly capturing the characteristics of the pdf (Pajevic and Basser, 2003). Objective Bayesian analysis (Behrens et al., 2003) is another approach that has been used to compute the pdf of DTI parameters, though this can be computationally much more demanding. Now that bootstrap can be performed even without repeated acquisitions, it will be interesting to compare bootstrap and Bayesian approaches in certain situations such as probabilistic fiber tracking using the same dataset.

Inadequacy of the single tensor DTI model to describe the data (as assumed by the design matrix used in the WLS fit to obtain the tensor) has been shown here to increase the errors of the bootstrap estimates. The model-based wild and residual bootstrap methods are particularly sensitive to this effect. The wild bootstrap method as implemented here is based on the symmetry of the probability distribution function (pdf) of the residuals. This assumption is violated by both low SNR data (due to log Rician noise) and single tensor assumptions. The residual bootstrap method is based on the similarity of the probability distribution functions between data points (and not the symmetry of the pdfs). However, depending on the alignment of the gradients relative to the tensor principal direction, the pdfs may vary among the data points due to log Rician noise and multi-tensor effects. Although also affected by the low SNR, the residual bootstrap method is less affected since it is not sensitive to the asymmetric pdfs due to log Rician noise, but only the differential effects on the pdfs among the gradient directions. Despite these effects, our results suggest that the model-based approaches generally perform better than the repetition-based methods. Model-based effects are not an issue when only 6 directions are used since this effectively reduces the diffusion ODF (orientation distribution function) (Tuch, 2004) to an exact effective single tensor.

Rician noise and PVE causing inadequacy of the single tensor model are important sources of uncertainty in the DTI derived parameters but there are other sources including cardiac pulsation, head motion, artifacts, eddy currents, magnetic susceptibility effects, etc. (Basser and Jones, 2002). The repetition-based methods are likely to better characterize the uncertainty from non-ideal variance caused by these sources than the model-based methods since repetition-based methods make less assumptions than model-based methods, though whether this holds true needs to be evaluated in some way. In this study, only ideal noise and PVE were considered because it is relatively straightforward to simulate these effects, but more studies are needed to evaluate how bootstrap methods perform with other sources of variance as well, either by simulating some aspects of these sources or by using real data.

Weighted linear least squares estimation was used for calculating the diffusion tensor from the original as well as bootstrap samples, but alternative ways to estimate the diffusion tensor exist. Nonlinear least squares estimator was shown to be more robust at high b values or low SNR (Jones and Basser, 2004) and was less likely to produce unphysical negative eigenvalues (Koay et al., 2006) than linear least squares. Furthermore, a robust estimator was shown to be effective against artifacts producing outliers (Chang et al., 2005; Mangin et al., 2002). Bootstrap can be combined with these estimators as well, though the additional computation time required by these processing methods instead of the computationally efficient linear least squares solution can be a limitation. For instance, computation time for nonlinear least squares estimation can be up to 60 times more than that of linear estimation (Chang et al., 2005). Considering the fact that bootstrap requires the tensor estimation to be iterated hundreds to thousands of times, whether it is beneficial to bootstrap with these more sophisticated tensor estimation and how bootstrap can be implemented more efficiently will be a subject of future study.

In summary, we have shown that a bias is present in the currently used repetition bootstrap method and have presented an alternate method (repetition bootknife) that corrects for this bias and, therefore, better estimates the standard errors of DTI parameters. We have also evaluated the model-based wild bootstrap which performs better than the repetition methods but is susceptible to model failures. We also present another model-based method (residual bootstrap) that generally performs better than all the other methods but is also sensitive to failures of the tensor model to describe the data. These results can be used to design DTI experiments in terms of choosing number of averages and number of diffusion sensitizing gradient directions to achieve the standard errors that permit observation of a particular effect sizes. Furthermore, importantly, the model-based methods enable probabilistic fiber tracking and hypothesis testing in longitudinal voxel-

Please cite this article as: SungWon Chung et al., Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.001.

wise analysis with a single acquisition, which allows maximization of the number of diffusion sensitizing directions in a clinically feasible scan time.

Acknowledgments

We are grateful to Jeffrey Berman, John Kornak, Pratik Mukherjee, and Michael Sdika for helpful discussions. This study was supported by RG3240A1 from the National Multiple Sclerosis Society.

References

Alexander, A.L., Hasan, K.M., Lazar, M., Tsuruda, J.S., Parker, D.L., 2001. Analysis of partial volume effects in diffusion-tensor MRI. Magn. Reson. Med. 45, 770–780.

Basser, P.J., Jones, D.K., 2002. Diffusion-tensor MRI: theory, experimental design and data analysis—A technical review. NMR Biomed. 15, 456–467.

Basser, P.J., Mattiello, J., LeBihan, D., 1994a. Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson., B 103, 247–254.

Basser, P.J., Mattiello, J., LeBihan, D., 1994b. MR diffusion tensor spectroscopy and imaging. Biophys. J. 66, 259–267.

Behrens, T.E., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., et al., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50, 1077–1088.

Chang, L.C., Jones, D.K., Pierpaoli, C., 2005. RESTORE: robust estimation of tensors by outlier rejection. Magn. Reson. Med. 53, 1088–1095.

Davidson, R., Flachaire, E., 2001. The Wild Bootstrap, Tamed at Last. Queen's University, Department of Economics, Working Papers.

Davison, A.C., Hinkley, D.V., 2003. Bootstrap Methods and their Application. Cambridge Univ. Press, Cambridge, UK.

Efron, B., 1979. 1977 Rietz lecture—Bootstrap methods—Another look at the jackknife. Ann. Stat. 7, 1–26.

Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.

Hasan, K.M., Alexander, A.L., Narayana, P.A., 2004. Does fractional anisotropy have better noise immunity characteristics than relative anisotropy in diffusion tensor MRI? An analytical approach. Magn. Reson. Med. 51, 413–417.

Heim, S., Hahn, K., Samann, P.G., Fahrmeir, L., Auer, D.P., 2004. Assessing DTI data quality using bootstrap analysis. Magn. Reson. Med. 52, 582–589.

Hesterberg, T.C., 2004. Unbiasing the bootstrap–bootknife sampling vs. smoothing. section on statistics and the environment. Am. Stat. Assoc. 2924–2930.

Jones, D.K., 2003. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. Magn. Reson. Med. 49, 7–12.

Jones, D.K., 2004. The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study. Magn. Reson. Med. 51, 807–815.

Jones, D.K., Basser, P.J., 2004. "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. Magn. Reson. Med. 52, 979–993.

Jones, D.K., Pierpaoli, C., 2005. Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach. Magn. Reson. Med. 53, 1143–1149.

Koay, C.G., Carew, J.D., Alexander, A.L., Basser, P.J., Meyerand, M.E., 2006. Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging. Magn. Reson. Med. 55, 930–936.

Lazar, M., Alexander, A.L., 2005. Bootstrap white matter tractography (BOOT-TRAC). NeuroImage 24, 524–532.

Liu, R.Y., 1988. Bootstrap procedures under some non-iid models. Ann. Stat. 16, 1696–1708.

Mangin, J.F., Poupon, C., Clark, C., Le Bihan, D., Bloch, I., 2002. Distortion correction and robust tensor estimation for MR diffusion imaging. Med. Image Anal. 6, 191–198.

O'Gorman, R.L., Jones, D.K., 2005. How many bootstraps make a buckle? 13th Annual Meeting of ISMRM. ISMRM, Miami Beach, p. 225.

Pajevic, S., Basser, P.J., 2003. Parametric and non-parametric statistical analysis of DT-MRI data. J. Magn. Reson. 161, 1–14.

Pierpaoli, C., Basser, P.J., 1996. Toward a quantitative assessment of diffusion anisotropy. Magn. Reson. Med. 36, 893–906.

Rao, J.N.K., Wu, C.F.J., 1988. Resampling inference with complex survey data. J. Am. Stat. Assoc. 83, 231–241.

Schwartzman, A., Dougherty, R.F., Taylor, J.E., 2005. Cross-subject comparison of principal diffusion direction maps. Magn. Reson. Med. 53, 1423–1431.

Shao, J., 1996. Resampling methods in sample surveys. Statistics 27, 203–254.

Shao, J., 2003. Impact of the bootstrap on sample surveys. Stat. Sci. 18, 191–198.

Tuch, D.S., 2004. Q-ball imaging. Magn. Reson. Med. 52, 1358–1372.

Whitcher, B., Tuch, D.S., Wang, L., 2005. The wild bootstrap to quantify variability in diffusion tensor MRI. 13th Annual Meeting of ISMRM. ISMRM, Miami Beach, p. 1333.

Please cite this article as: SungWon Chung et al., Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.001.