Supplemental Results:

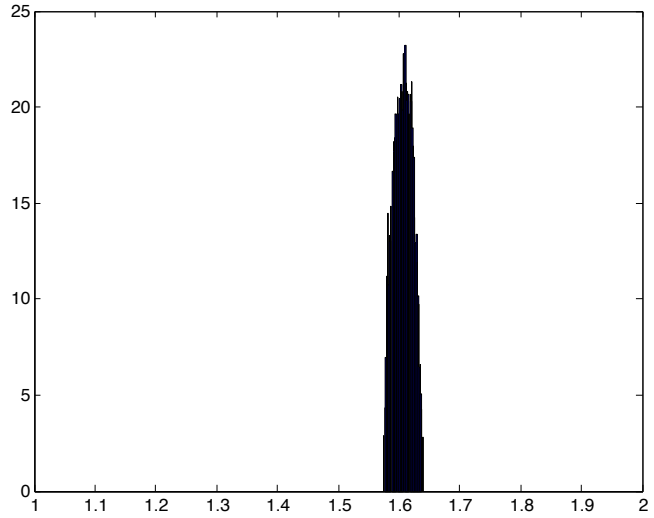
DTI encoding direction condition number

Below, we provide the metrics for the condition number of the PNC 64 direction encoding scheme in the table and figure below (Table is adapted from Batchelor et al., 2003). As can be seen the mean condition number of the b-matrix is comparable to several popular and widely used encoding schemes (e.g. Jones 30) and is superior to others tested in Batchelor et al., 2003.

Supplemental Table 1:

Condition number metrics for popular DTI direction schemes				
Scheme	κ_{max}	κ_{mean}	κ_{min}	σ_{κ}
PNC 64	**1.570**	**1.606**	**1.639**	**0.015**
Jones 6	1.584	1.582	1.579	0.001
Jones 20	1.619	1.599	1.563	0.013
Jones 30	1.598	1.590	1.576	0.004
DSM6	2.000	1.883	1.341	0.131
DSM20	2.000	1.883	1.341	0.131
DSM30	2.000	1.883	1.341	0.131
Tetra	9.229	7.431	5.850	0.831
Dual-gradient	2.000	1.852	1.605	0.097
Tetra-ortho	1.620	1.604	1.528	0.020
Icosahedral	1.581	1.581	1.581	0
Note: Table adapted from Batchelor et al., 2003				

Supplemental Figure 1. Frequency histogram of condition numbers for 10^4 uniformly distributed rotations.



ROI specific DTI metrics

Supplemental Table 2. Mean and standard deviations of FA and MD values for data failing QA (Poor data) and data passing QA (Good+Excellent).

FA	ATR_L	ATR_R	CST_L	CST_R	CGC_L	CGC_R	CGH_L	CGH_R	Forceps major	Forceps minor	IFO_L	IFO_R	ILF_L	ILF_R	SLF_L	SLF_R	UF_L	UF_R
Fail QA (Poor)	0.359 (0.024)	0.342 (0.024)	0.529 (0.027)	0.533 (0.027)	0.493 (0.058)	0.455 (0.062)	0.366 (0.046)	0.350 (0.042)	0.564 (0.037)	0.422 (0.027)	0.416 (0.024)	0.424 (0.025)	0.435 (0.025)	0.463 (0.029)	0.346 (0.023)	0.361 (0.026)	0.401 (0.034)	0.406 (0.036)
Pass QA (Good +Excellent)	0.367 (0.019)	0.351 (0.017)	0.543 (0.020)	0.548 (0.018)	0.517 (0.044)	0.487 (0.058)	0.387 (0.033)	0.374 (0.34)	0.578 (0.029)	0.430 (0.018)	0.428 (0.020)	0.435 (0.20)	0.445 (0.019)	0.475 (0.024)	0.356 (0.018)	0.372 (0.022)	0.410 (0.024)	0.418 (0.025)
MD (cm^2/sec)	**ATR_L**	**ATR_R**	**CST_L**	**CST_R**	**CGC_L**	**CGC_R**	**CGH_L**	**CGH_R**	**Forceps major**	**Forceps minor**	**IFO_L**	**IFO_R**	**ILF_L**	**ILF_R**	**SLF_L**	**SLF_R**	**UF_L**	**UF_R**
Fail QA (Poor)	2.275 (0.113)	2.029 (0.128)	2.005 (0.101)	2.117 (0.096)	2.205 (0.134)	2.184 (0.152)	2.298 (0.167)	2.200 (0.184)	2.447 (0.323)	2.414 (0.130)	2.294 (0.110)	2.261 (0.104)	2.258 (0.128)	2.205 (0.125)	2.261 (0.106)	2.205 (0.119)	2.281 (0.172)	2.349 (0.147)
Pass QA (Good +Excellent)	2.262 (0.075)	2.034 (0.085)	2.011 (0.074)	2.113 (0.069)	2.209 (0.096)	2.154 (0.101)	2.260 (0.113)	2.162 (0.114)	2.460 (0.314)	2.419 (0.093)	2.278 (0.075)	2.253 (0.078)	2.244 (0.094)	2.192 (0.095)	2.246 (0.081)	2.190 (0.092)	2.293 (0.094)	2.346 (0.078)

Power Analysis

Using the association between age and FA reported in the manuscript we perform a power analysis to estimate the change in sample size necessary based on the inclusion or exclusion of poor data. All power estimate were computed using G*power (http://www.gpower.hhu.de/en.html) (Faul, 2007)

Power analysis indicates (α=0.05; 1-β=0.80) that 80% more individuals are needed if the effect size between age and FA is estimated from data that has not undergone rigorous QA (Table 1). Not surprisingly, significantly fewer samples are needed when using QAed data. This analysis fits within the range of sample sizes that are feasible number at single center studies. However, we would argue that these sample size estimates should incorporate a minimum of 10% data loss, particularly in sample including children and young adults. In additional, other work in DTI quality assurance provides detailed information on power analyses with local rejection of data (Lauzon et al., 2013).

Supplemental Table 3

Group	Effect Size/Correlation	a	Power (1-b)	Estimated Sample Size
Passed QA	0.51	0.05	0.80	26
All (includes Failed QA)	0.41	0.05	0.80	47*

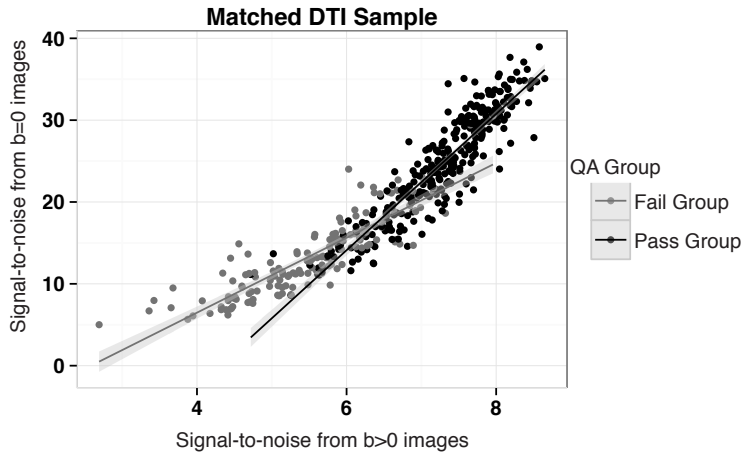
Signal-to-noise ratio as a DTI QA metric
Use of b>0 TSNR vs. b=0 TSNR

Our TSNR measurement is atypical for DTI. Typically, SNR is estimated from the non-weighted images (b=0). Therefore, we computed TSNR across the 7 b=0 images for all subjects. Overall, we find that TSNR from b=0 and TSNR from b>0 are highly correlated (r=0.90) in the full sample. Correlations between b=0 TSNR and b>0 remain high when estimated in data passing (r=0.91) and failing manual QA (r=0.88; Supplemental Figure 2A).

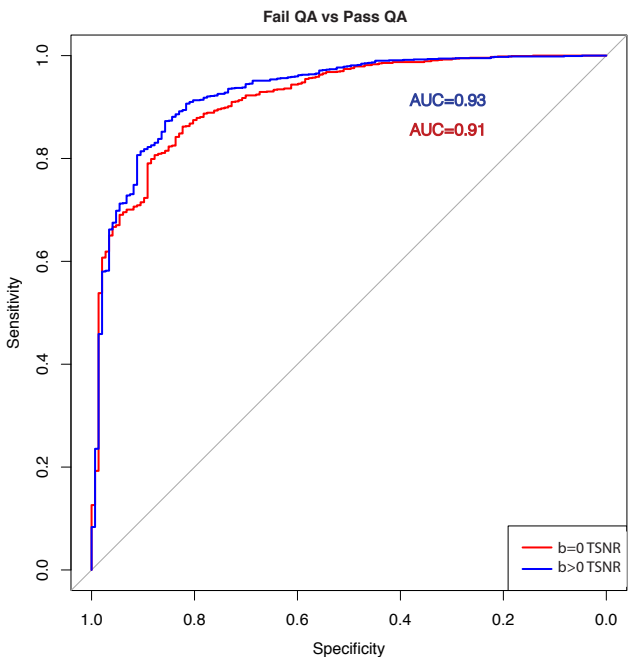
Since we propose using TSNR as a cut-off for data quality we also estimated the between subject variance in each method. We find between subject variance to be much higher for the b=0 TSNR approach as opposed to the b>0. Estimating TSNR from more volumes may explain why b>0 have lower between. In addition, the coefficient of variation was higher using the b=0 approach (61%) than the b≠0 approach (13%). Thus, we would argue that TSNR from the 64 weighted images provides a more representative, and potentially more usable (at least for QA) estimate of SNR. Not surprisingly, substituting the more typical TSNR measure in our ROC analysis resulted in poorer overall prediction; however, this prediction was still quite high (~91%; Supplement Figure 2B).

Supplemental Figure 2
A. The correlation between SNR derived from non-weighted images (b=0) and weighted images (b>0) was high (r=0.91). **B.** Substituting the more typical TSNR measure (b=0) in the ROC analysis resulted in poorer overall prediction; however, this prediction was still quite high (91% vs. 93%).

A



B



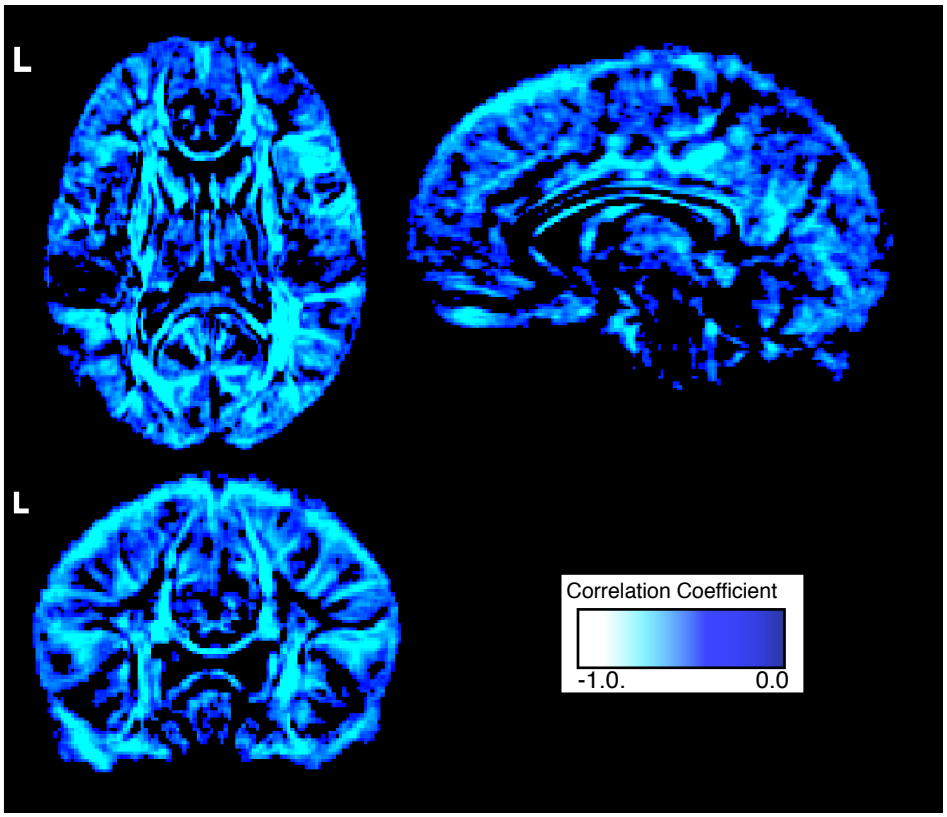
Association between TSNR and FA

Here, we show that there is a negative association between anisotropy and TSNR (Supplemental Figure 3). Thus, high FA regions (e.g. corpus callosum) have the lowest TSNR. TSNR is likely lower in areas with the highest FA since signal is dependent on the direction measured. Given that we sample many different directions there is likely to more variability in TSNR from image to image where anisotropy is high. Thus TSNR, as a whole

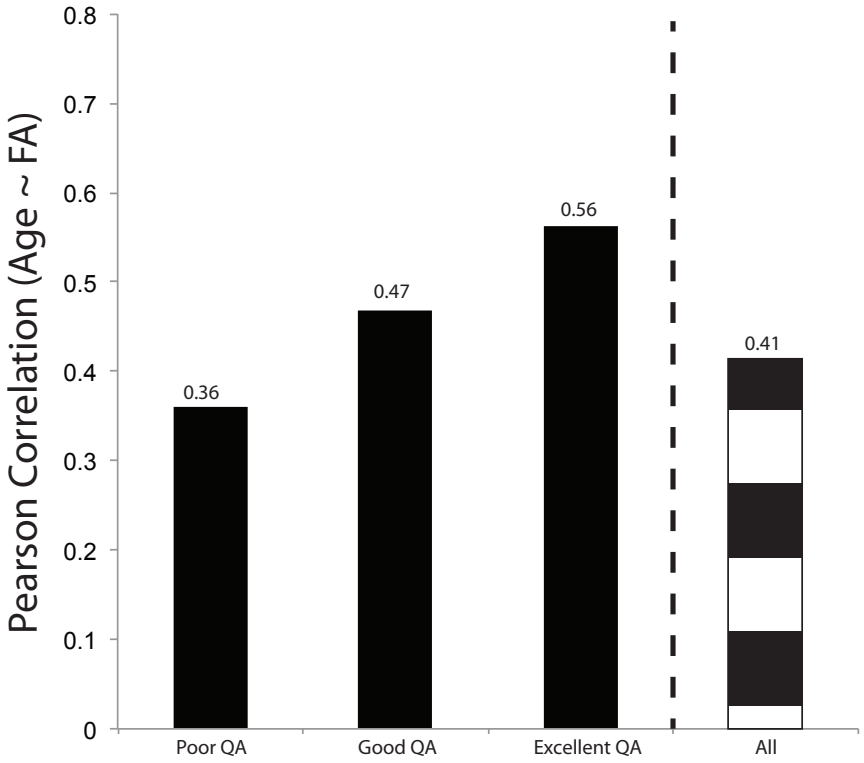
brain metric is useful for QA, however, its utility in highly anisotropic regions of the brain may be limited.

However, TSNR differs between the QA groups on a regional basis. While not as specific as the voxelwise analysis, quantitative measurement of TSNR indicated significant differences in all GM ROIs (Harvard-Oxford Atlas), on average, and in most, but not all white matter ROIs (JHU White Matter Tracts). Those individuals that pass QA had higher TSNR values than those failing QA in all GM and WM regions except in the corticospinal tracts CCST). We now provide these quantitative plots in Figure 6 (means +/- standard deviation), but note that voxelwise analysis indicates that areas of high anisotropy are negatively associated with TSNR. Not surprisingly, the effect sizes for mean difference in TSNR were much higher in GM (Cohen's d range: 1.23-2.07) than WM (Cohen's d range: (0.01-1.72).

Supplemental Figure 3: Regions of high anisotropy have the lowest TSNR.
Negative association between FA and TSNR. This is most prominent in in white matter tracts where anisotropic diffusion is highest (e.g. corpus callosum).



Supplemental Figure 4
Pearson correlation coefficients between Age and FA in the three manual QA groups (n=146 each). Poor data had a significantly lower correlation between age and FA as compared to data in the Excellent QA group ($p<.05$). Data in the Good QA group fell in between the other two groups. The striped bar shows Pearson correlation when all data is combined. Estimating this association is strongest when data of low quality is excluded from analysis. MD showed the same pattern, but with negative correlations.



References
Batchelor, P.G, Atkinson, D., Hill, D.L.G., Calamante, F., & Connelly, A. (2003). Anisotropic noise propagation in diffusion tensor MRI sampling schemes. Magnetic Resonance in Medicine, 49: 1143-1151.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39:* 175-191.