



RESEARCH ARTICLE

Fast quality control method for derived diffusion Metrics (YTTRIUM) in big data analysis: U.K. Biobank 18,608 example

Ivan I. Maximov^{1,2,3}  | Dennis van der Meer^{2,4} | Ann-Marie G. de Lange^{1,2,5,6} | Tobias Kaufmann² | Alexey Shadrin² | Oleksandr Frei^{2,7} | Thomas Wolfers^{1,2} | Lars T. Westlye^{1,2,8} 

¹Department of Psychology, University of Oslo, Oslo, Norway

²Norwegian Centre for Mental Disorders Research (NORMENT), Department of Mental Health and Addiction, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway

³Department of Health and Functioning, Western Norway University of Applied Sciences, Bergen, Norway

⁴School of Mental Health and Neuroscience, Faculty of Health Medicine and Life Sciences, Maastricht University, The Netherlands

⁵LREN, Centre for Research in Neurosciences – Department of Clinical Neurosciences, CHUV and University of Lausanne, Lausanne, Switzerland

⁶Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK

⁷Center for Bioinformatics, Department of Informatics, University of Oslo, Oslo, Norway

⁸KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway

Correspondence

Ivan I. Maximov, NORMENT, Department of Psychology, University of Oslo, Oslo, Norway.
Email: ivan.maximov@psykologi.uio.no

Funding information

Norges Forskningsråd, Grant/Award Number: 249795

Abstract

Deriving reliable information about the structural and functional architecture of the brain in vivo is critical for the clinical and basic neurosciences. In the new era of large population-based datasets, when multiple brain imaging modalities and contrasts are combined in order to reveal latent brain structural patterns and associations with genetic, demographic and clinical information, automated and stringent quality control (QC) procedures are important. Diffusion magnetic resonance imaging (dMRI) is a fertile imaging technique for probing and visualising brain tissue microstructure in vivo, and has been included in most standard imaging protocols in large-scale studies. Due to its sensitivity to subject motion and technical artefacts, automated QC procedures prior to scalar diffusion metrics estimation are required in order to minimise the influence of noise and artefacts. However, the QC procedures performed on raw diffusion data cannot guarantee an absence of distorted maps among the derived diffusion metrics. Thus, robust and efficient QC methods for diffusion scalar metrics are needed. Here, we introduce Fast quality control method for derived diffusion Metrics (YTTRIUM), a computationally efficient QC method utilising structural similarity to evaluate diffusion map quality and mean diffusion metrics. As an example, we applied YTTRIUM in the context of tract-based spatial statistics to assess associations between age and kurtosis imaging and white matter tract integrity maps in U.K. Biobank data ($n = 18,608$). To assess the influence of outliers on results obtained using machine learning (ML) approaches, we tested the effects of applying YTTRIUM on brain age prediction. We demonstrated that the proposed QC pipeline represents an efficient approach for identifying poor quality datasets and artefacts and increases the accuracy of ML based brain age prediction.

KEYWORDS

brain maturation, diffusion QC, DKI, DTI, U.K. Biobank, WMTI, YTTRIUM

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Diffusion magnetic resonance imaging (dMRI) provides a range of structural brain features based on routine clinical measurements, which has contributed to its popularity across fields and applications (de Lange et al., 2020; Kochunov et al., 2015; Westlye et al., 2010). Advanced dMRI is technically challenging and often involves time-consuming acquisitions placing high demands on the performance and stability of the scanner hardware. Therefore, dMRI data are vulnerable to experimental setup perturbations including post-processing approaches, which might bias the results. In turn, optimised post-processing pipelines (Ades-Aron et al., 2018; Maximov, Alnæs, & Westlye, 2019; Tournier et al. 2019) and stringent procedures for quality control (QC; Alfaro-Almagro et al., 2018; Bastiani et al., 2019; Graham, Drobnyak, & Zhang, 2018; Haddad et al., 2019) are important to increase reliability and sensitivity. Various approaches have been developed to detect and correct artefacts in raw diffusion data originating, for example, from eddy currents, bulk head motions, susceptibility distortions (Andersson & Sotiropoulos, 2016), noise (Kochunov et al., 2018), Gibbs ringing artefacts (Perrone et al., 2016; Veraart, Fieremans, Jolescu, Knoll, & Novikov, 2016; Veraart, Novikov, et al., 2016), presence of outliers (Koch, Zhukov, Stöcker, Groeschel, & Schultz, 2019) and diffusion metric variability (David, Mesri, Viergever, & Leemans, 2019; Maximov et al., 2015).

However, QC and data harmonisation procedures applied on raw diffusion data (Fortin et al., 2017; Mirzaalian et al., 2018) do not guarantee accurate numerical computation of scalar diffusion metrics. Derived diffusion metrics from diffusion or kurtosis tensors are sensitive to a range of subject-specific factors such as age or various brain disorders, but also to applied numerical algorithm or its programming implementation (David et al., 2019; Grinberg et al., 2017; Lebel et al., 2012; Maximov et al., 2015). The effects of noisy observations on subsequent between-subjects analysis involving the derived diffusion metrics can be mitigated using simple outlier detection procedures (see, e.g., de Lange et al., 2020; Richard et al., 2018; Tønnesen et al., 2018). However, few publications have directly assessed the effects of QC filtration of final data and performing a sanity check of the derived scalar maps before the statistical analysis. As an example, one can use a visual inspection (see, e.g., *slicesdir* utility from FSL [Smith et al., 2007]) or truncation based on variability of the data and their *SD*. We know that outliers might affect the results of analysis, in particular, machine learning (ML) algorithms and related prediction or classification output. One example is brain age prediction using neuroimaging data (Kaufmann et al., 2019; Smith, Vidaurre, Alfaro-Almagro, Nichols, & Miller, 2019), where corrupted data either in the training or test sets will influence the accuracy of the prediction.

Here, we introduce a QC method for the derived diffusion maps based on twofold parameterisation: first, diffusion data reduction based on the scalar diffusion values averaged across skeleton voxels using tract-based spatial statistics (TBSS; Smith et al., 2007), and, second, structural similarity (SSIM; Wang, Bovik, Sheikh, & Simoncelli, 2004) of individual diffusion maps relative to the mean diffusion image derived from all subjects. We demonstrate feasibility of this approach for U.K. Biobank (UKB) data (Miller et al., 2016) using

three commonly applied diffusion approaches: diffusion tensor imaging (DTI) (Basser, Mattiello, & Lebihan, 1994), diffusion kurtosis imaging (DKI) (Jensen, Helpert, Ramani, Lu, & Kaczynski, 2005) and white matter tract integrity (WMTI; Fieremans, Jensen, & Helpert, 2011). We evaluated the effect of the developed QC approach by assessing age-diffusion associations and the accuracy of brain age prediction using ML technique.

2 | METHODS AND MATERIALS

2.1 | Participants and MRI data

We used dMRI data obtained from 18,608 subjects (see Figure 1 for age and sex distribution). An accurate overview of the UKB imaging acquisition parameters and initial QC pipeline can be found in Alfaro-Almagro et al. (2018) and Miller et al. (2016). Briefly, a conventional Stejskal-Tanner monopolar spin-echo echo-planar imaging (EPI) sequence was used with multiband factor 3, diffusion weightings (*b*-values) were 1 and 2 ms/μm² and 50 non-coplanar diffusion directions per shell. All subjects were scanned at 3T Siemens Skyra scanners with a standard Siemens 32-channel head coil, in Cheadle and Newcastle, U.K. The spatial resolution was 2 mm³ isotropic, and five AP versus three PA images with *b* = 0 ms/μm² were acquired. All diffusion data were post-processed using an optimised diffusion pipeline (Maximov et al., 2019) consisting of six steps: noise correction (Veraart, Fieremans, et al., 2016; Veraart, Novikov, et al., 2016), Gibbs-ringing correction (Kellner, Dhital, Kiselev, & Reiser, 2016), estimation of echo-planar imaging distortions, head motions, eddy-current and susceptibility distortions (Andersson & Sotiropoulos, 2016), spatial smoothing using *fslmaths* from FSL (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012) with a 1 mm³ Gaussian kernel, and diffusion metrics estimation using Matlab scripts (MathWorks, Natick, MA; Veraart, Sijbers, Sunaert, Leemans, & Jeurissen, 2013). UKB data were processed using the high-performance computing facility Colossus

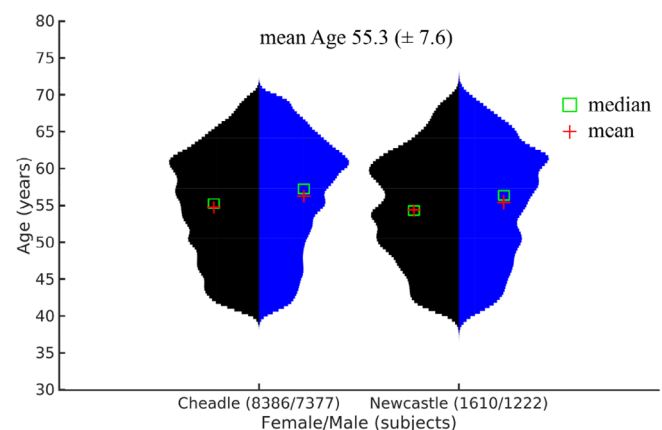


FIGURE 1 Demographic data depending on the scanner location and gender. The mean age (*SD*) for all data is on the top of the plot

at the University of Oslo and large data storage located at Services for Sensitive Data (TSD).

2.2 | Diffusion metrics

2.2.1 | DTI and DKI

Diffusion signal decay can be represented as the Taylor expansion along diffusion weightings (Novikov, Kiselev, & Jespersen, 2018). This can be approximated by two diffusion tensors of the second (DTI) and fourth (DKI) orders of diffusion wavevector. A set of scalar maps are derived from eigenvalues of the both tensors such as FA, mean, axial and radial diffusivities (MD, AD, RD, respectively), and mean, axial and radial kurtosis (MK, AK and RK, respectively). The scalar maps characterise integrative features of brain tissue with potential to represent sensitive biomarkers (Jones, 2010).

2.2.2 | WMTI

In the frame of standard diffusion model (Novikov, Kiselev, & Jespersen, 2018), WMTI represents an intra-axonal space as a bundle of cylinders with effective radius equals to zero (Fieremans et al., 2011). The cylinders are impermeable, that is, there is no water exchange between intra- and extra-axonal spaces. The extra-axonal space is described by anisotropic Gaussian diffusion. In order to keep the model simple a few more assumptions have been made: intra-axonal space consists of mostly myelinated axons without any contribution from myelin due to a fast relaxation rate across of typical diffusion times; at the same time in extra-axonal space the glial cells possess fast water exchange with extra-cellular matrix; both intra- and extra-axonal spaces are modelled by Gaussian diffusion tensors. In order to avoid degeneration (Jelescu, Veraart, Fieremans, & Novikov, 2016), intra-axonal diffusion is assumed to be slower than diffusion in extra-axonal matrix. However, this assumption should be considered carefully, because it artificially reduces a set of plausible estimations appearing in the conventional diffusion experiments (Novikov, Kiselev, & Jespersen 2018; Veraart, Novikov, & Fieremans, 2018; Novikov, Veraart et al., 2018). Besides, WMTI parameterisation works in the case of quite coherent axonal bundle with an orientation dispersion below 30° (Fieremans et al., 2011). WMTI allows one to derive axonal water fraction (AWF), extra-axonal axial and radial diffusivities (axEAD and radEAD, respectively).

2.2.3 | Tract based spatial statistics

In order to evaluate and compare different QC approaches for the derived diffusion maps, we applied TBSS (Smith et al., 2007). Initially, all FA volumes were aligned to the FMR158_FA template, supplied by FSL, using non-linear transformation implemented by FNIRT (Andersson & Jenkinson, 2019). Next, a mean FA image across 18,600 subjects was

obtained and thinned in order to create mean FA skeleton. Afterwards, each subject's FA data are projected onto the mean FA skeleton, by filling the skeleton with FA values from the nearest relevant tract centre. TBSS minimises confounding effects due to partial voluming and residual misalignments originated from non-linear spatial transformations. For each diffusion metric, we computed the individual skeleton projecting the non-FA values onto the FA skeleton.

2.2.4 | QC model description

Our approach of image quality estimation originates from multi-dimensional experiments in nuclear magnetic resonance spectroscopy (Ernst, Bodenhausen, & Wokaun, 1987), when an additional dimension allows one to resolve hidden resonance peaks. Natural parameter in diffusion scalar metrics is an absolute value which either has its physical limitations, for example, FA and AWF lie between 0 and 1 and diffusion kurtosis is limited by [0, 3] range (Tabesh, Jensen, Ardekani, & Helpert, 2010; Veraart, Van Hecke, & Sijbers, 2011), or some region-specific values from other sources, for example, free water diffusivity in the brain equals to 3 $\mu\text{m}^2/\text{ms}$. Thus, by applying a reasonable threshold rule one can discard volumes with unfeasible values from further analysis. However, since the values are typically averaged over the volume (region of interest, skeleton, etc.), we still can expect “hidden” outliers with minimal influence on the averaged metric.

The general workflow of the proposed QC algorithm is summarised in Figure 2. The workflow consists of five steps: first, we estimate diffusion metrics for each subject in the diffusion space; second, a normalisation step is performed in order to align each diffusion map to Montreal Neurological Institute (MNI) space using FA map and derived non-linear transformation; third, two QC parameters are estimated for each subject: namely, averaged diffusion metric and SSIM. SSIM values are estimated using the cohort mean diffusion metric as a reference, mean diffusion values are obtained by averaging the scalar maps over the TBSS skeleton; fourth, k-means approach for one cluster allows us to obtain a distribution of the Euclidean distances for each subject point to the cluster centroid; finally, the median distance and empirically determined number of neighbours are used for the density based clustering in order to identify possible outliers among the derived diffusion metrics. As a result, the outlier exclusion is done in the level of the whole brain volume.

We assume that SSIM allows us to spread image parameterisation into the second dimension using three principal features: luminance, contrast and structure as following (Wang et al., 2004),

$$f_{ssim} = \left(\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \right)^\alpha \left(\frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \right)^\beta \left(\frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \right)^\gamma,$$

where index x belongs to the evaluated map and y to the population mean (reference) map for the given diffusion metric, $\mu_{x,y}$ are the means of x and y , $\sigma_{x,y}^2$ are the variances of x and y and σ_{xy} is the covariance of x and y , constants $c_{1,2,3}$ are the variables stabilising the

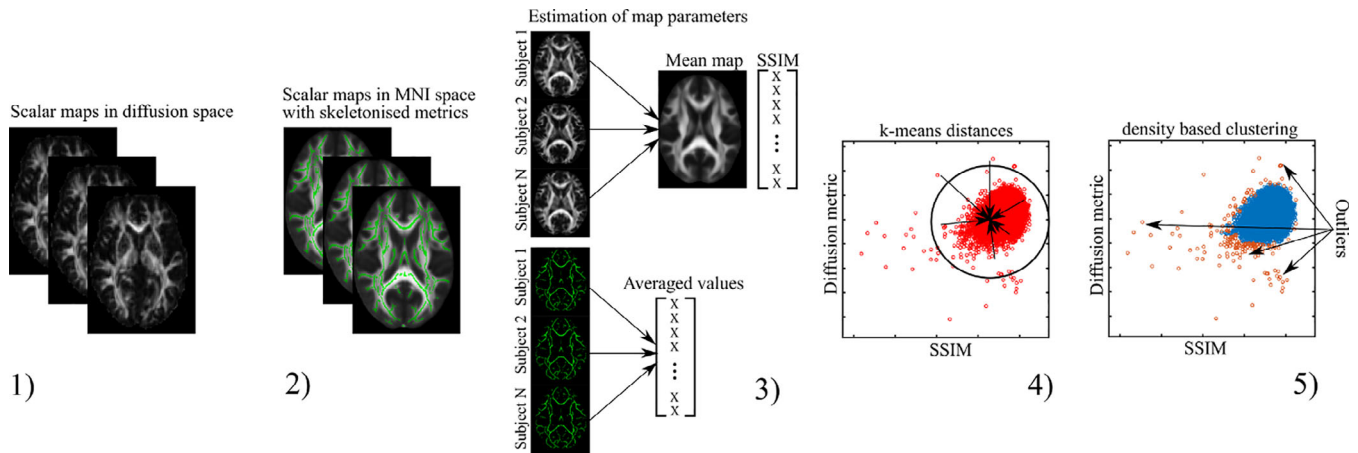


FIGURE 2 Algorithmic workflow of the developed QC procedure for diffusion metric. The proposed procedure consists of five steps: (1) estimation of diffusion scalar maps; (2) transferring of scalar maps into TBSS format using conventional scheme (Smith et al., 2007); (3) estimation of SSIM and skeleton-averaged values for each subject. As a reference image for the SSIM estimation, we used a mean diffusion map; (4) applying of k-means for estimation of all point distances to the cluster centroid; (5) data filtration using the density-based spatial clusterisation

SSIM estimation, and α , β and γ are the weights of three SSIM features. The stabilisation constants c_1 , c_2 and c_3 are defined in Matlab as $c_1 = (0.01 \cdot L)^2$; $c_2 = (0.03 \cdot L)^2$; $c_3 = c_2/2$; with L specified by a dynamic range value: $L = 1$ for images with $[0,1]$ scale, and 255 for others. By design, the SSIM metric is devoted to extract structural differences between the images, in contrast to the conventional approaches based on pixelwise error visualisation. Thus, SSIM is capable to identify differences between the information extracted from the scalar reference image and target image, similar to human visual perception. For the estimations, we used the *ssim* function, implemented in Matlab. The SSIM values are estimated for each diffusion metric separately, such as FA, MD, MK and so on, due to different map structure of the diffusion metrics. In theory, SSIM estimator can be improved for some specific purposes (Charrier, Knoblauch, Maloney, Bovik, & Moorthy, 2012) or generalised (Brunet, Vrscay, & Wang, 2012). Nevertheless, original SSIM metric already proved its capability in medical image quality verification (Chow & Paramesran, 2016; Renieblas, Nogués, González, Gómez-Leon, & del Castillo, 2017; Vinding et al., 2017). The weights α , β and γ allow one to emphasise the principle SSIM features in order to enhance a contrast between original image and reference. While SSIM weights adjustment is still debated (Li, & Bovik, 2009), we empirically define the following weights $\alpha = .1$, $\beta = .1$ and $\gamma = 2$, in order to stretch a range of SSIM value components.

After the diffusion metric evaluation and normalisation of the scalar maps to the common space, we estimated the SSIM metrics for each diffusion map using averaged normalised diffusion metric as a reference image in SSIM. We performed outlier detection using the following two-step approach: first, we used *k*-means clustering (implemented as *kmeans* Matlab function) to define one cluster based on squared Euclidian distances for (diffusion metric, SSIM) pairs (David & Vassilvitskii, 2007). Next, in order to introduce object density parameterisation, we used the median distance of the distance

distribution (MDD) around the cluster centroid as a unit for neighbourhood radius in density-based scan algorithm with noise (*dbscan*, implemented as MATLAB function; Daszykowski, Walczak, & Massart, 2002; Ester, Kriegel, Sander, & Xu, 1996). The optional parameters in the *dbscan* algorithm are number of objects in neighbourhood of a central object and number of MDD units. We empirically set it to be equal to 10 and 7, respectively, in the tests. As such, the chosen parameters allow us to apply cluster density estimation independently from the original data distribution (see, e.g., Figure 3 and Kendall correlation coefficients between diffusion metrics and SSIM values, estimated using Matlab function *corr*).

As a frequently applied QC approach for comparison purpose, we applied a simple threshold approach of $|3| SD$ from the mean diffusion value after regressing out main effects of age, sex and site.

2.3 | Statistical analysis

In order to assess the effects of our proposed QC pipeline on the sensitivity of the diffusion metrics we tested for associations with age and sex using linear models as implemented in the Matlab function *lmfit*. In the subsample of 799 subjects (724 UKB subjects + 75 artificially distorted maps) we employed the following general linear model (GLM): $y = b_0 + b_1 \text{Age} + b_2 \text{Sex} + b_3 \text{Site}$, where Age is given in years, Sex and site as a dichotomous variable. We computed specificity and sensitivity of the automated artefact correction. Sensitivity is defined as a ratio of True Positive/(True Positive + False Negative) and specificity as a ratio of True Negative/(True Negative + False Positive).

In the full UKB sample we employed the following models: $y = b_0 + b_1 \text{Age} + b_2 \text{Sex} + b_3 \text{Site} + (b_4 \text{Age}^2)$. We computed root mean squared error (RMSE) and R-squared as proxies for goodness-of-fit. We compared coefficients between models (before and after discarding datasets flagged by our QC pipeline) using the R package *cocor*

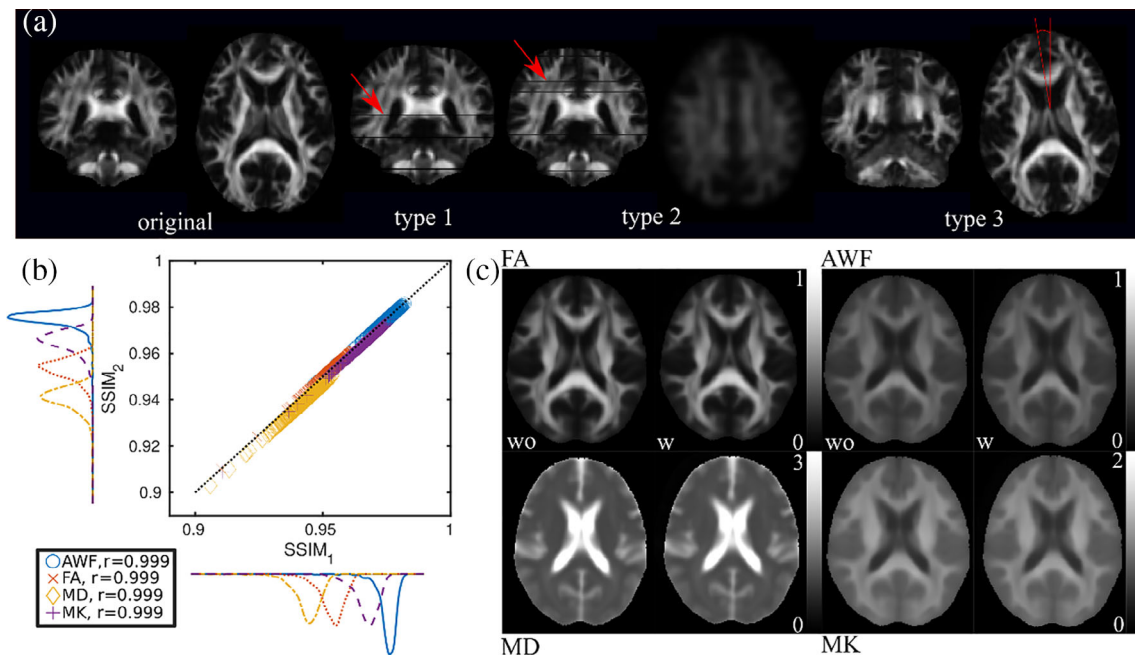


FIGURE 3 Example of simulated distortions added to the evaluated diffusion scalar maps and their influence on the estimated SSIM values. (a) an original FA image and three types of distortions: Type 1—random zero slices (see the red arrow), Type 2—slices with scaled and smoothed values (see the red arrow), Type 3—rotation of a whole volume around the superior–inferior axis (red lines show the angle); (b) correlation of the SSIM metrics evaluated for FA, AWF, MD and MK metrics using data without (SSIM₁) and with (SSIM₂) outliers. r is the Pearson correlation coefficient; black dotted line is a unity line; (c) left-side images are mean metrics averaged without (wo) outliers, when right-side images are mean metrics including (w) the outliers into averaging step

(Diedenhofen & Musch, 2015). In order to assess normality of the residuals from the linear models we used QQ-plots (Aldor-Noiman, Brown, Buja, Rolke, & Stine, 2013; implemented as *qqplot* Matlab function) and Kolmogorov–Smirnov (KS) test with W critical value (Kolmogoroff, 1933; Smirnov, 1948). The W critical values have been used as indirect measures of normality of the residuals. The KS tests were implemented as MATLAB function *kstest*.

2.4 | Machine learning for brain age gap estimation

We estimated the influence of outliers on ML based brain age prediction. The brain age gap (BAG) is defined as the difference between chronological and predicted age and has been proposed to reflect a sensitive imaging-derived phenotype (Kaufmann et al., 2019). For age prediction we applied two frequently used approaches. First, we employed linear model and multiple regressors (LMMR) defined as $Y = X\beta - \delta$, where Y is the chronological age, β is the regressor vector, X is the matrix of brain features used for prediction, and δ is the BAG. The solution can be obtained by pseudo-inversion X^+ matrix (Smith et al., 2019). In order to improve the ML-training, we used 25% of eigenstates produced by the singular value decomposition replacing the X matrix as recommended by Smith et al., 2019. The estimations were performed using original Matlab script from Smith et al., 2019 (<http://www.fmrib.ox.ac.uk/BrainAgeDelta>) with removed cross-validation code. The second algorithm is XGBoost employing a

gradient boosting approach (Chen & Guestrin, 2016). The extreme gradient boosting algorithm has been shown to have enhanced performance and speed in sequential decision trees ML algorithms. The XGBoost algorithm allows one to use an optimised loss function instead of increased weights as well as regularisation multipliers. The parameters of XGBoost were chosen as follows: eta = 1; number of rounds = 250; max depth = 4; lambda = 10^{-7} . Estimations were performed using Julia (<https://julialang.org>) implementation of XGBoost algorithm (<https://github.com/dmlc/XGBoost.jl>).

For simplicity, we used the following linear model using only four diffusion metrics averaged over the skeleton: $Y \sim b_1 \text{FA} + b_2 \text{MD} + b_3 \text{MK} + b_4 \text{AWF} + b_5 \text{Sex} + b_6 \text{Site}$. In order to assess the influence of outliers on age prediction, a fixed number of 476 outliers, identified by the proposed QC approach over all diffusion metrics, was combined with varying samples of good-quality data, creating total training-sets of 1,000, 2,000, 3,000, 4,000, 5,000, 7,500, 10,000, 12,500 and 15,000 subjects. The 476 outliers were manually added to each sample, leading to outlier percentages of 47.6, 23.8, 15.9, 11.9, 9.52, 6.35, 4.76, 3.81 and 3.17%, respectively. In all training sets, we kept the sex and site distribution identical. All training sets were selected from the whole UKB dataset. Thousand subjects not included in the training sets were selected as a test sample that was used in all runs. We performed the BAG estimations separately for the training sets with and without outliers, respectively. As criteria we used the Pearson correlations between chronological and predicted ages, and root mean squared errors estimated for the test sample.

2.5 | Simulated artefacts in small subsample

In order to verify our approach for detection of possible “badly” estimated scalar metrics and outliers we used a random subset of UKB data consisting of 724 subjects. Next, we manually introduced three types of image distortions to the evaluated diffusion scalar maps in MNI space. The first type (Type 1) is based on complete loss of N_1 random slices in the image volume. In our case, we set upper bound of N_1 to be equal 5. The second type of artefacts (Type 2) is based on value scaling of up to $N_2 = 7$ random slices. Scaling of the random slices can be performed in two ways as a division or multiplication of the diffusion values. To dilute scaled values between neighbouring slices we applied 3D Gaussian smoothing with 3mm^3 kernel. The final type of distortions (Type 3) is based on residual misalignments along an image normalisation process. As a simple implementation of the residual misalignments we used rotation around superior–inferior axis with a random angle up to 5° . An example of original diffusion maps and three types of artificial distortions is presented in Figure 3. Finally, we added 25 volumes (10.4% of original data) of each type of distortions to four diffusion metrics: FA, MD, MK and AWF. In order to test influence of artefacts on the derived mean metric maps, we evaluated SSIM parameters of initial datasets using mean maps with (799 volumes) and without (724 volumes) outliers.

3 | RESULTS

3.1 | Sensitivity to simulated artefacts

In the randomly selected subset of UKB data we evaluated an influence of artificial outliers on the averaged diffusion metrics and estimated SSIM metrics. Resulting SSIM correlations are presented in Figure 3. High linear correlations (over .999) demonstrated that an introduction of outliers into data subsets did not influence on mean reference images and, consequently, on the SSIM evaluations. Supporting Information provide examples of the diffusion maps detected in the whole UKB sample with different types of distortions after data processing and scalar metric evaluation.

Figure 4 shows an application of the developed QC method to the subsample consisting of 799 subjects with artificially introduced outliers of the three types. The sensitivity of the QC method for the diffusion metrics is presented in Table 1. Briefly, based on AWF we detected all three types of introduced outliers. In the case of FA, we missed 10 (13%) outliers; in the case of MD, we missed 7 (9%) outliers; in the case of MK, we missed 2 (3%) outliers. In contrast, the QC approach based on data truncation beyond three SDs from the mean allowed us to detect for FA, only six outliers (69 outliers are missed, 92%); for MD, it detected only four outliers (71 outliers are missed, 95%); for MK, it detected six outliers (69 outliers are

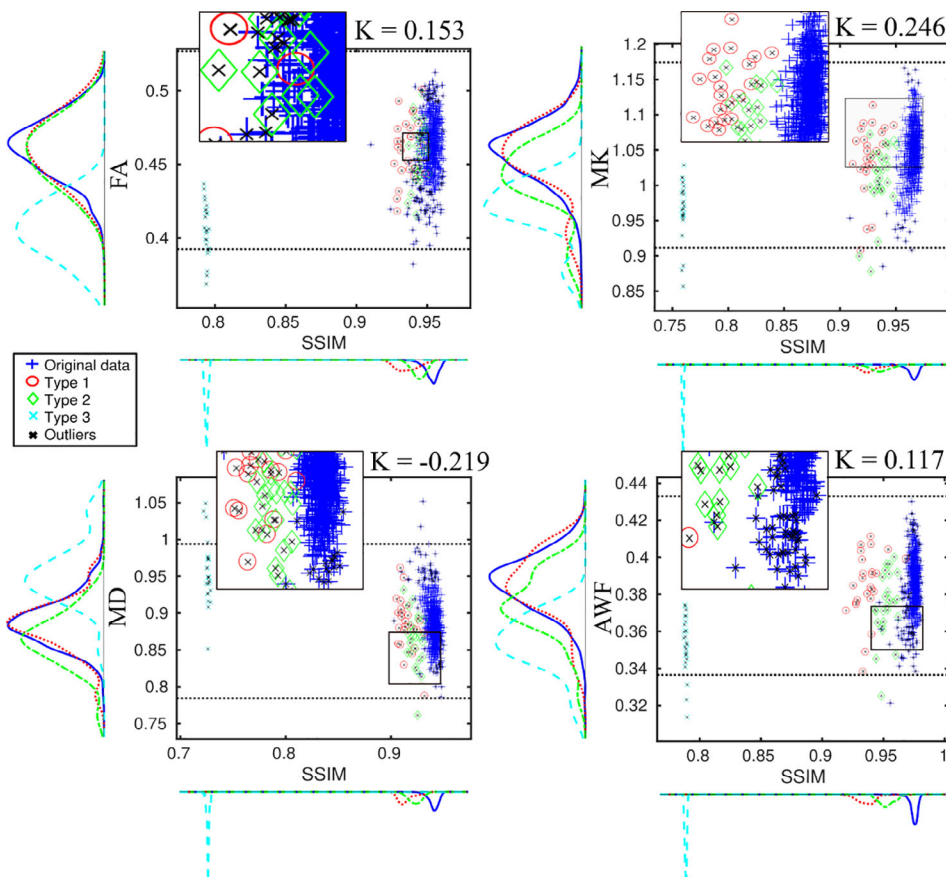
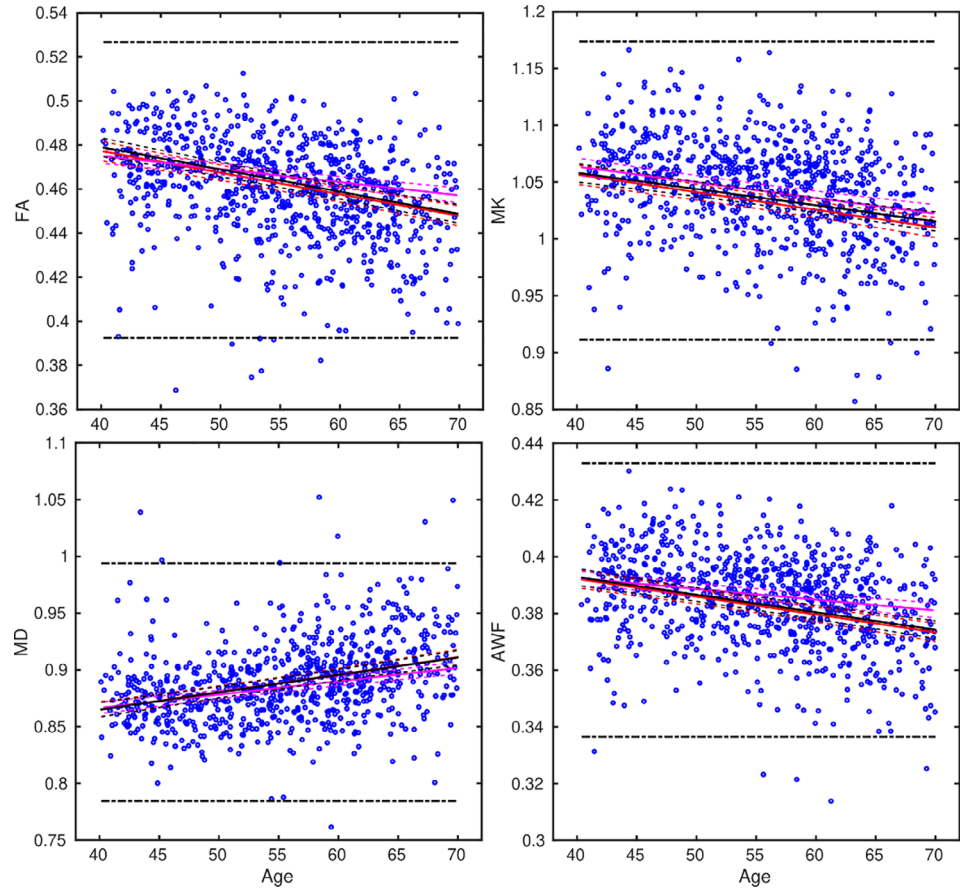


FIGURE 4 Scatter plots of diffusion metrics and SSIM values for the data with three types of outliers. All artificial outliers are marked by the different colours. The result of QC method is marked by black crosses. Image inserts demonstrate a zoomed boundary between outlier groups and original data. Estimated Kendall correlation coefficients (K) for the diffusion metrics and SSIM are presented as well. SSIM, FA, MK and AWF are unit-less values, MD is in $\mu\text{m}^2/\text{ms}$. The black dashed lines are boundaries of three SDs from the mean and delineate the simple QC method based on data thresholding

TABLE 1 Sensitivity and specificity of the proposed QC method for the diffusion metrics based on artificial distortions of three types (see Figure 3)

Metrics	Missed volumes from 25 outliers (sensitivity/specificity)			Number of discarded volumes. False positive
	Type 1	Type 2	Type 3	Original data
FA	2 (0.92/0.80)	8 (0.68/0.80)	0 (1/0.80)	142
MD	1 (0.96/0.91)	6 (0.76/0.91)	0 (1/0.91)	69
MK	0 (1/0.97)	2 (0.92/0.97)	0 (1/0.97)	21
AWF	0 (1/0.85)	0 (1/0.85)	0 (1/0.85)	112

FIGURE 5 Results of general linear model $y = b_0 + b_1 \text{ Age} + b_2 \text{ Sex} + b_3 \text{ Site}$ for four diffusion metrics over the test data sample of 724 subjects. The solid and dashed red lines are linear fit (LF) and the interval of confidence (CI 95%); the black solid and dashed lines are LF and CI for data thresholded by three SDs from the mean diffusion metrics (marked as dot-dash lines); the magenta solid and dashed lines are LF and CI for the proposed QC method. SSIM, FA, MK and AWF are unit-less values, MD is in $\mu\text{m}^2/\text{ms}$ 

missed, 92%); and for AWF, it detected four outliers (71 outliers are missed, 95%).

Figure 5 shows the various diffusion metrics plotted as a function of age and the corresponding linear fits based on the GLM. Briefly, the raw data and thresholding method yielded similar GLM parameters (see Table 2 for the intercept and slope values). Cocor function revealed no significant slope differences for any of the diffusion metrics between the raw and QC'ed data. Table 2 summarises the goodness-of-fit measures for the selected diffusion metrics for three datasets (raw data, thresholding QC and the developed QC method) and GLM parameters (b_0 , intercept, and b_1 , age slope). QQ plots and the W parameters from KS tests based on FA, MD, MK and AWF for three datasets (raw data, thresholding QC and our QC method)

suggested that our proposed QC method yields the most “normal” residuals (Figure 6).

3.2 | Effects of QC pipeline on the sensitivity to age, sex and scanner site

Figure 7 shows an application of the developed QC method to the UKB data with 18,608 subjects. As detailed above, we discarded datasets defined as outliers based on mean skeleton diffusion metrics and SSIM. The mean diffusion maps used as a reference for SSIM estimations are depicted in Supporting Information. Distributions of relevant diffusion metrics and demographics of the data

TABLE 2 Results of GLM $y = b_0 + b_1 \text{ Age} + b_2 \text{ Sex} + b_3 \text{ Site}$ for four diffusion metrics using test sample of 724 subjects

Metrics/statistics	Raw data No 799		Threshold with 3 SD			Our QC method		
	RMSE	R ²	NO	RMSE	R ²	NO	RMSE	R ²
FA	0.0211	.112	792	0.0198	.137	592	0.0135	.1
MD	0.0327	.125	786	0.03	.141	657	0.0223	.144
MK	0.0416	.0999	791	0.0391	.0985	705	0.0345	.105
AWF	0.0152	.103	793	0.0145	.108	611	0.0109	.0629
	Intercept	Slope		Intercept	Slope		Intercept	Slope
FA	0.5164	-9.80·10 ⁻⁴		0.5197	-1.01·10 ⁻³		0.4990	-5.98·10 ⁻⁴
MD	0.8037	1.54·10 ⁻³		0.8037	1.53·10 ⁻³		0.8194	1.18·10 ⁻³
MK	1.1188	-1.55·10 ⁻³		1.1154	-1.43·10 ⁻³		1.1189	-1.38·10 ⁻³
AWF	0.4174	-6.29·10 ⁻⁴		0.4176	-6.19·10 ⁻⁴		0.4077	-3.80·10 ⁻⁴

Note: RMSE is the root mean squared error; R² is the R-squared parameter; NO is the number of observations; SDs; Intercept is b₀; Slope is b₁.

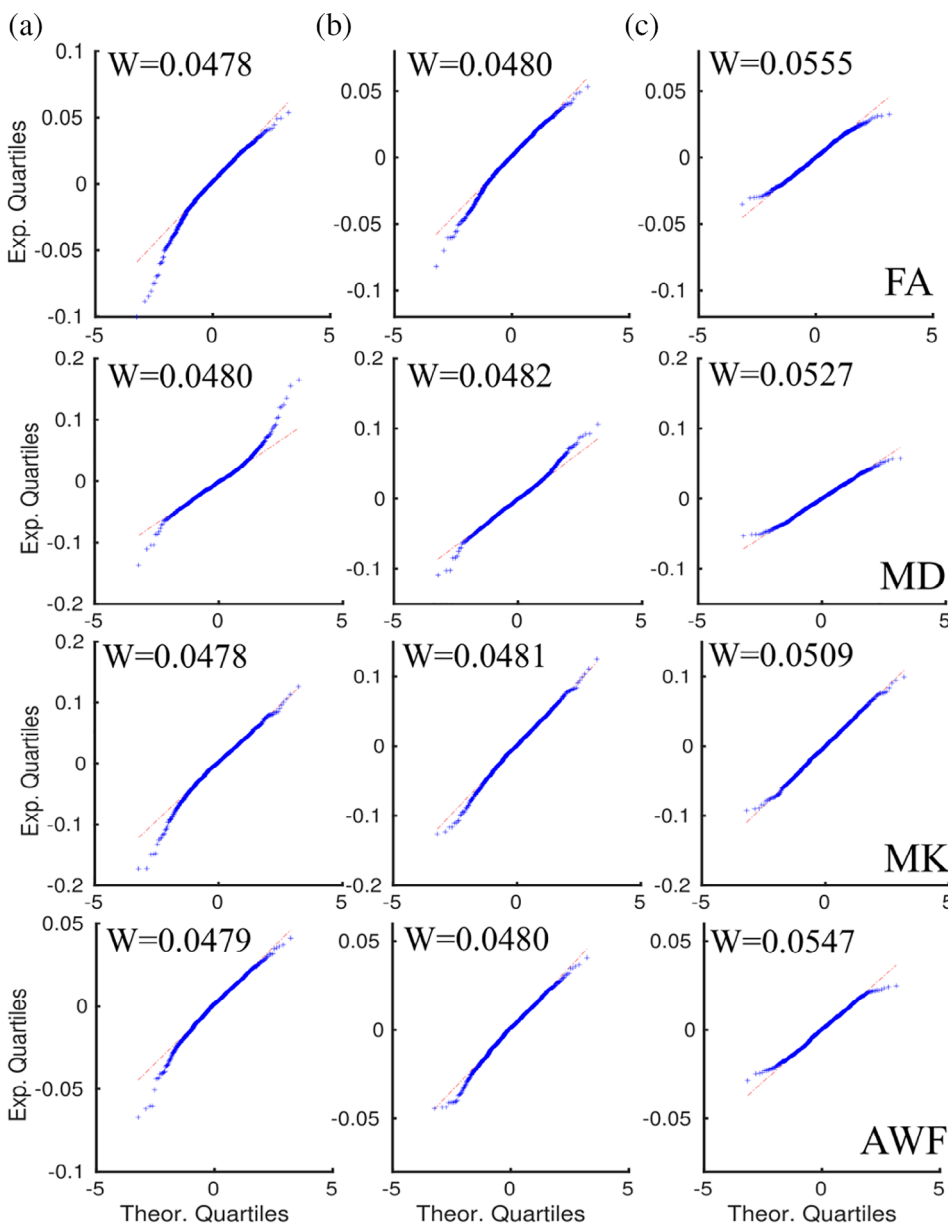


FIGURE 6 QQ plots of the GLM residuals for three cases (see Figure 4): (a) raw data; (b) after thresholding by three SDs; (c) the proposed QC method. Values W are the critical numbers of Kolmogorov–Smirnov (KS) test for a normality. In all cases, KS test did not reveal that the residuals are normally distributed

FIGURE 7 An application of QC method to UKB data. Mean diffusion maps are presented as a reference for the SSIM estimations. The red circles are identified outliers; the blue circles are the filtered data. SSIM, FA, MK and AWF are unit-less values, MD is in $\mu\text{m}^2/\text{ms}$. The dashed black lines mark the boundaries in three SDs from the mean value

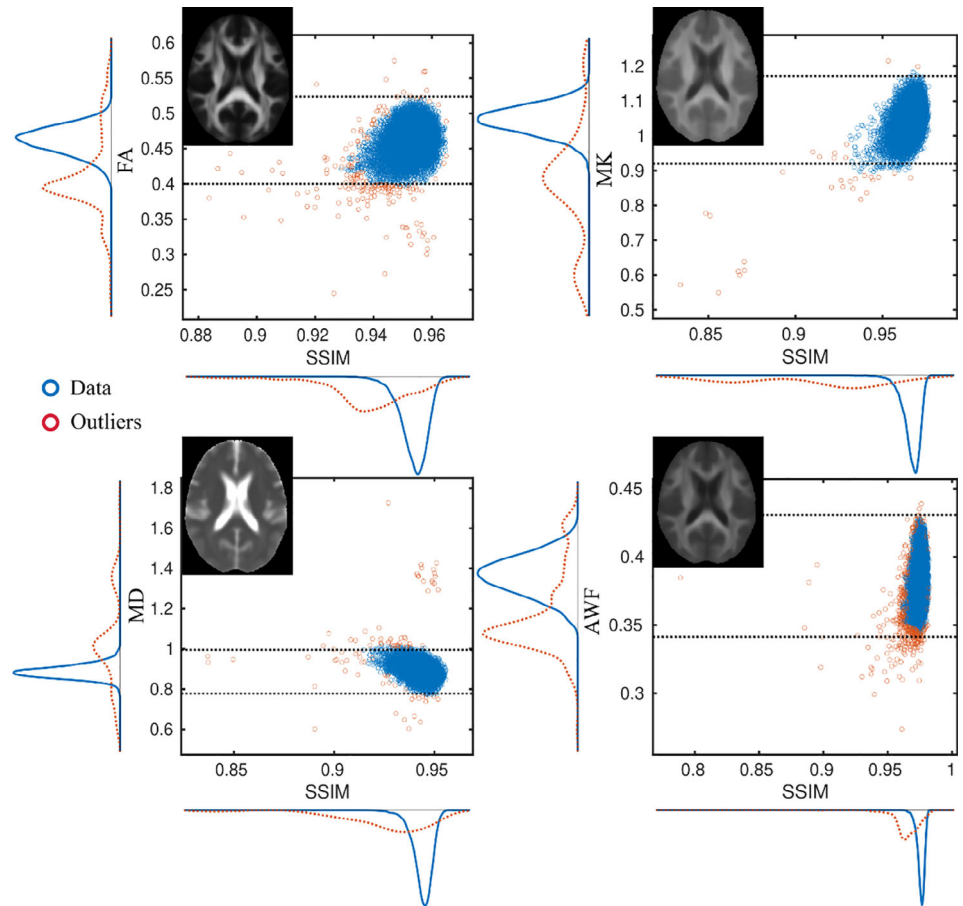
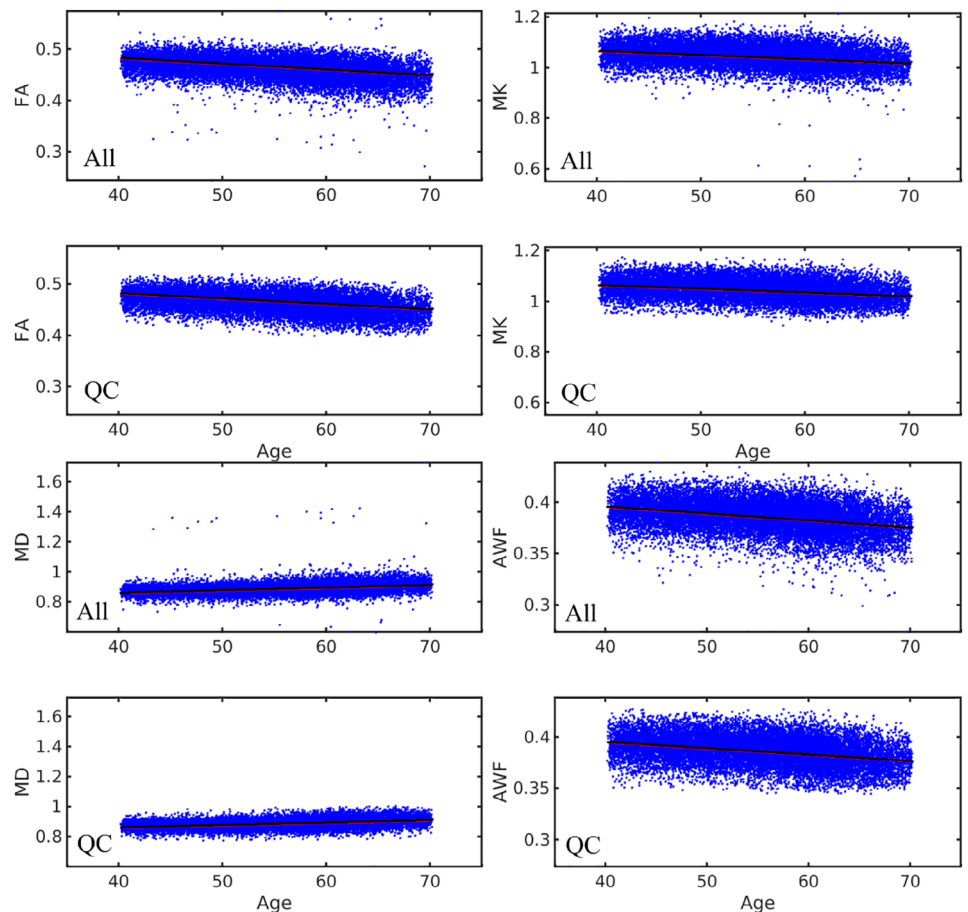


FIGURE 8 The results of GLM age-diffusion correlations with linear age term (red line) and quadratic age term (black line). The plots marked as “All” consists of all raw data ($n = 18,608$); the plots marked as “QC” consists of data passed through the QC filtration ($n = 18,132$). Intervals of confidence (CI 95%) are presented as dashed line in all cases. FA, MK and AWF are unit-less values, MD is in $\mu\text{m}^2/\text{ms}$



Metrics/statistics	All data, linear age term			QC passed, linear age term		
	NO	RMSE	R ²	NO	RMSE	R ²
FA	18,608	0.0187	.173	18,134	0.0171	.179
MD	18,603	0.0337	.133	18,129	0.0282	.158
MK	18,608	0.0398	.1	18,134	0.037	.096
AWF	18,607	0.014	.119	18,133	0.0131	.113
		Intercept	Slope		Intercept	Slope
FA		0.5277	-1.12·10 ⁻³		0.5242	-1.05·10 ⁻³
MD		0.7914	1.70·10 ⁻³		0.7976	1.58·10 ⁻³
MK		1.1315	-1.64·10 ⁻³		1.1231	-1.47·10 ⁻³
AWF		0.4226	-6.70·10 ⁻⁴		0.4196	-6.10·10 ⁻⁴
Metrics/statistics	All data, quadratic age term			QC passed, quadratic age term		
	NO	RMSE	R ²	NO	RMSE	R ²
FA	18,608	0.0187	.176	18,134	0.0171	.182
MD	18,603	0.0335	.141	18,129	0.028	.165
MK	18,608	0.0397	.106	18,134	0.0369	.101
AWF	18,607	0.014	.125	18,133	0.0131	.117
		Intercept	Slope		Intercept	Slope
FA		0.5283	-1.13·10 ⁻³		0.5247	-1.06·10 ⁻³
MD		0.7898	1.73·10 ⁻³		0.7962	1.61·10 ⁻³
MK		1.1332	1.67·10 ⁻³		1.1246	-1.50·10 ⁻³
AWF		0.4232	-6.81·10 ⁻⁴		0.4201	-6.19·10 ⁻⁴

Note: RMSE is the root mean squared error; R² is the R-squared parameter, NO is the number of observations; Intercept is b_0 ; Slope is b_1 .

defined as outliers are presented in Supporting Information. A higher number of outliers were identified from the Cheadle ($n = 396$; 3%) site compared to the Newcastle ($n = 78$; 1%) site, and both sex (39% of women) and age distribution (58.55/7.74 years) of the outlier data did not diverge substantially from the distributions in the total sample. For illustration, we depicted the boundaries of $|3|$ SD from the mean cohort values for each diffusion metric.

Figure 8 presents age-related trajectories (linear and quadratic fits) for four diffusion metrics with the detected outliers included or excluded. The summary statistics are summarised in Table 3. All other metrics are shown in the Supporting Information. Figure 9 shows corresponding QQ-plots of the GLM residuals in the case of linear and quadratic age terms. QQ-plots of the residuals for all diffusion metrics are shown in Supporting Information. Briefly, the residuals from the models using raw data show strong deviations from the diagonal in both linear and quadratic age terms. In contrast, the residuals from QC approved data appear more normal. The GLM age-diffusion dependence and related QQ-plots for simple thresholding approach are presented in Supporting Information. In short, the results for truncation QC method repeat the same behaviour as in the case of simulated data (see Figures 5 and 6). Cocor function revealed no significant slope differences for any of the diffusion metrics between the raw and QC'ed data in Figure 8.

TABLE 3 Results of two GLM $y = b_0 + b_1 \text{ Age} + b_2 \text{ Sex} + b_3 \text{ Site} + (b_4 \text{ Age}^2)$ for four diffusion metrics using all UKB data with and without QC procedure

3.3 | Effects of QC pipeline on the ML BAG estimations

Figure 10 shows the trajectories of RMSE and correlations between chronological and predicted ages for the two ML algorithms. For statistics and cross validation of the BAG results, we repeated model training 100 times, randomly choosing the training samples from whole UKB data. Briefly, in the case of QC filtered data model performance increased only moderately with sample size, with RMSE of the XGBoost algorithm suggesting only minor effects. In contrast, the training sets with outliers demonstrated strong dependence of the chronological and predicted age correlations and RMSE on the percentage of outliers in the training sample, in particular, for the LMMR approach. For both ML algorithms, increasing training sample size decreased RMSE in the test set.

4 | DISCUSSION

Advanced dMRI offers sensitive measures of brain tissue micro- and macrostructural architecture and integrity, with large potential for the basic and clinical neurosciences. With the surge of large-scale clinical and population-based efforts acquiring dMRI data from thousands of individuals, there is an increasing need to develop computationally

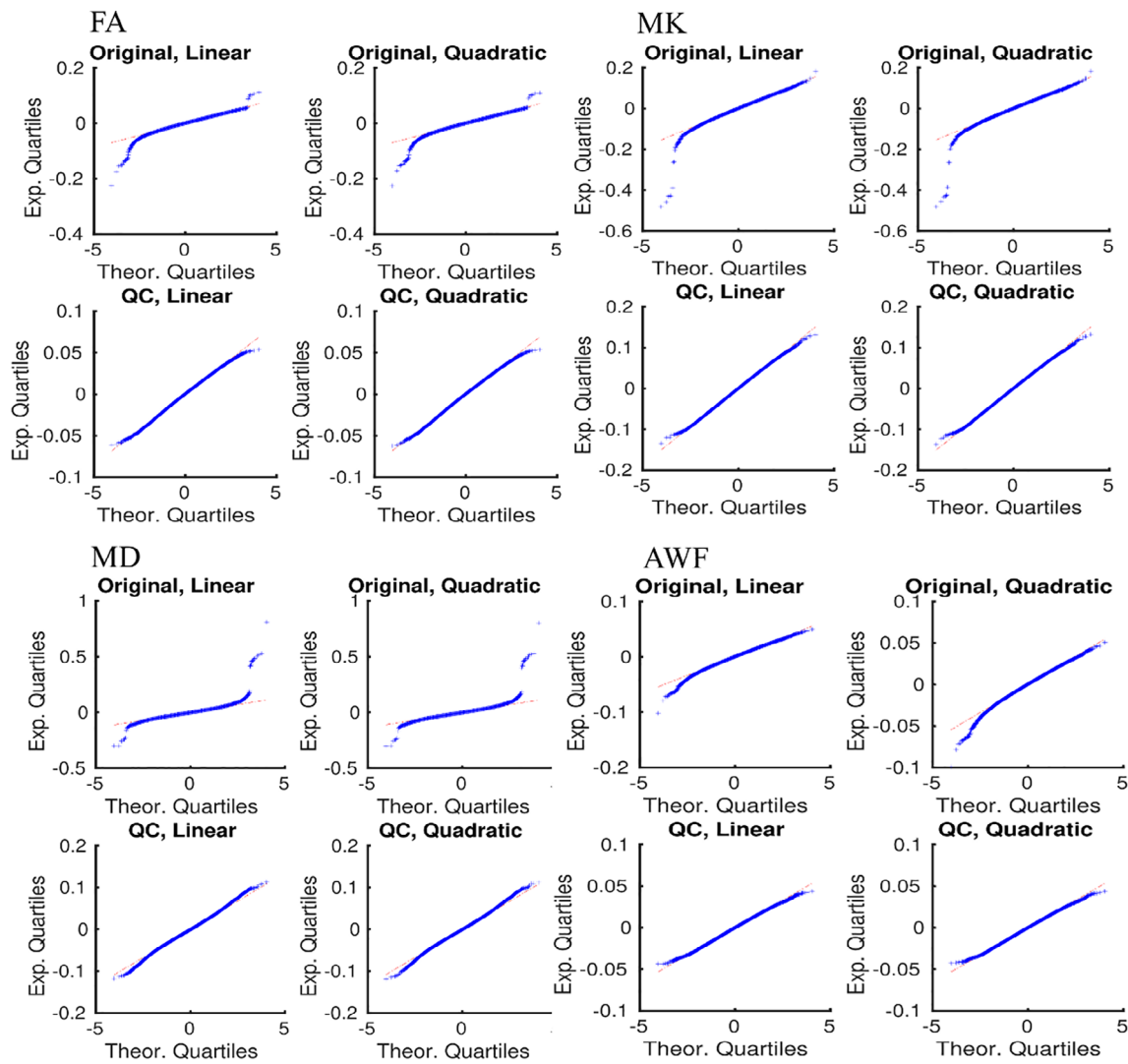


FIGURE 9 QQ-plots of the GLM residuals for four diffusion metrics. “Original, Linear” means the all data and GLM with linear age term; “Original, Quadratic” means the all data and GLM with quadratic age term; “QC, Linear” means the QC filtered data and GLM with linear age term; “QC, Quadratic” means the QC filtered data and GLM with quadratic age term

efficient pipelines for quality assessment and identification of poor quality data among derived diffusion scalar metrics. The proposed QC method based on 2D data representation exploiting similarity metrics and data density features enables an efficient evaluation of data quality after estimation of diffusion scalar metrics. In a subsample ($n = 724$ plus 75 artificial outliers), our semi-automated artefact detection based on similarity metrics yielded high sensitivity and specificity and the residuals from linear age fits in the full sample ($n = 18,608$) resulted in more normal residuals after compared to before discarding flagged datasets. Additionally, the QC pipeline improved brain-age prediction using ML by mitigating the influence of outliers in the training set.

By default, the harmonised validated raw diffusion data allow one to derive accurate scalar metrics. However, a quality evaluation of the processed diffusion maps is still an open question in big data analysis. Many efforts have been made to develop accurate QC and harmonisation procedures on raw diffusion weighted data (Fortin

et al., 2017; Mirzaalian et al., 2018). Nevertheless, derived diffusion metrics from DTI or DKI may still deviate from expected range, for example, due to remaining artefacts and numerical misestimations (see Supporting Information for examples of the distorted diffusion maps). Despite improved post-processing algorithms (Ades-Aron et al., 2018) for raw diffusion data, there is no consensus yet about a unified pipeline for diffusion data, for example, noise correction methods are regularly revised (Muckley et al., 2021), Gibbs ringing artefacts can remain in the images due to different origins such as a partial Fourier (Muckley et al., 2021), frequency drift effect (Vos et al., 2016) can bias the estimations, in particular in the case of advanced dMRI protocols, and diffusion gradient non-linearity correction (Rudrapatna, Parker, Roberts, & Jones, 2020) might be important as well. Notably, a number of artefacts in the scalar diffusion maps could be minimised by applying a state-of-the-art algorithms such as, for example, eddy_gpu, if a computational facility allows that. The simple considerations of 2D representations of the averaged diffusion

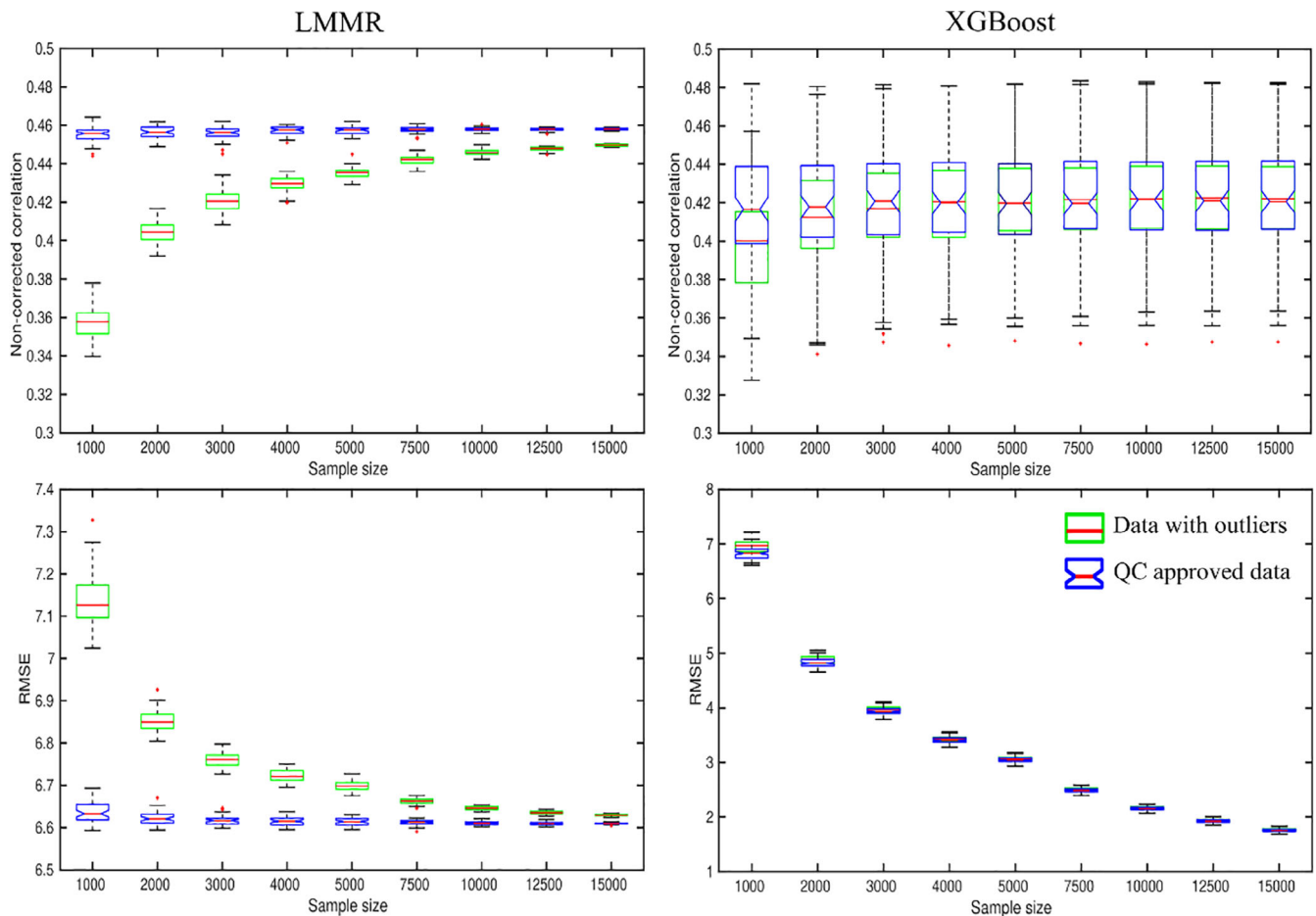


FIGURE 10 Outlier influence on the brain age predictions for two ML algorithm: linear model with multiple regressors (LMMR) and gradient boosting method (XGBoost). The top row is non-corrected Pearson correlations between chronological and predicted ages as a function of the sample size of training sets. The correlations were estimated for the fixed test sample of 1,000 subjects. The bottom row is root mean square error (RMSE) of the predictions as a function of the sample size of training sets. The rectangular green boxplots are the datasets with included outliers, the blue boxplots with “notched” feature are the datasets without outliers (QC filtered)

metric and SSIM values are an advantage of the developed QC method. This allows us to take into account frequently applied measures in large-scale studies, that is, diffusion metrics averaged across a region of interest or the entire TBSS skeleton and the structural similarity based on the intensity, contrast and structure of the scalar map in relation to a reference map. As an improvement of the developed approach, one could generalise SSIM parameter for the scalar maps belonging to the same diffusion approach. A combination of various diffusion contrasts could improve the QC efficacy and reduce a computation time. Our simulations revealed that our method is capable of identifying image artefacts with different origins with high sensitivity and specificity. Notably, our approach allows one to reveal different artefacts originated either from the corrupted diffusion maps or caused by the not accurate warping procedure to MNI space. In the case of map misalignments, the original diffusion maps in the diffusion space still can be used in the native space, for example, for a tractography but should be processed separately in the following group analysis. This is particularly valuable in the context of large-scale studies, where manual QC is not feasible and when a

quantitative estimate of structural similarity is needed. Whereas our direct comparison of slopes did not reveal a significant effect of the QC procedure on the estimated age-associations, the linear models based on QC'ed data yielded evidence of improved model fits in terms of the distributions of the model residuals compared to models based on the non-QC'ed data.

We found evidence of improved ML based age prediction when limiting the number of noisy datasets in the training set. In general, larger training sets are expected to increase accuracy of brain age prediction (see, e.g., Kaufmann et al., 2019). However, in practice, the number of accessible data is usually limited. Thus, it is very important to know how different amounts of undetected outliers in the training set could affect the prediction accuracy in an independent test set. Our results demonstrated that a higher portion of outliers in the training set influenced the prediction accuracy in the test set. Surprisingly, for XGBoost, in contrast to the RMSE, the correlation between predicted and chronological age did not increase much with increasing training set size. For LMMR the correlation coefficients increased in accordance with increased sample size. In both instances, however,

the proportion of bad datasets in the training set influenced the prediction accuracy in the test set, with markedly improved prediction with lower proportion of noisy data.

In summary, although an overall beneficial effect of removing poor quality datasets results is not surprising, our results serve as relevant demonstrations of the importance of QC in the context of large-scale studies. It should also be noted that all datasets included in the current analysis have been checked and approved by the initial U.K. Biobank QC procedures (Alfaro-Almagro et al., 2018), and the reported effects of noise removal on age-associations are likely to represent lower-bound effects compared to a scenario with no initial QC procedures. In general, whereas minimising noise is a universal aim, the direct effects and value of QC will vary between studies and applications. As a relevant verification of the proposed QC approach, we plan to apply the same procedure to other imaging biobanks such as Adolescent Brain Cognitive Development study, conceived and funded by the National Institutes of Health, United States.

Conclusively, in the case of big data, automated, efficient and reliable approaches for evaluating the scalar diffusion metrics prior to statistical analysis are needed. Our results suggest that our proposed method is suitable as a complementary test of the estimated diffusion data to increase sensitivity of conventional diffusion scalar metrics.

ACKNOWLEDGEMENT

This work was funded by the Research Council of Norway (249795 and 276082). This research has been conducted using the U.K. Biobank under Application 27412. This work was performed on the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT). Computations were also performed on resources provided by UNINETT Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway. IIM thanks Dr. Viljami Sairanen for his help with FSL and *slurm* queue scripting.

DATA AVAILABILITY STATEMENT

All used data are accessible through U.K. Biobank service.

ORCID

Ivan I. Maximov  <https://orcid.org/0000-0001-6319-6774>

Lars T. Westlye  <https://orcid.org/0000-0001-8644-956X>

REFERENCES

- Ades-Aron, B., Veraart, J., Kochunov, P., McGuire, S., Sherman, P., Kellner, E., ... Fieremans, E. (2018). Evaluation of the accuracy and precision of the diffusion parameter Estimation with Gibbs and Noise removal pipeline. *NeuroImage*, 183, 532–543. <https://doi.org/10.1016/j.neuroimage.2018.07.066>
- Aldor-Noiman, S., Brown, L. D., Buja, A., Rolke, W., & Stine, R. A. (2013). The power to see: A new graphical test of normality. *The American Statistician*, 67(4), 249–260. <https://doi.org/10.1080/00031305.2013.847865>
- Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., ... Smith, S. M. (2018). Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*, 166(February), 400–424. <https://doi.org/10.1016/j.neuroimage.2017.10.034>
- Andersson, J., Jenkinson, M., & Smith, S. (2007). Non-linear registration aka spatial normalisation FMRIB technical report TR07JA2. Retrieved from <https://www.fmrib.ox.ac.uk/datasets/techrep/tr07ja2/tr07ja2.pdf>.
- Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, 125(January), 1063–1078. <https://doi.org/10.1016/j.neuroimage.2015.10.019>
- Basser, P. J., Mattiello, J., & Lebihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin Echo. *Journal of Magnetic Resonance, Series B*, 103(3), 247–254. <https://doi.org/10.1006/jmrb.1994.1037>
- Bastiani, M., Cottaar, M., Fitzgibbon, S. P., Suri, S., Alfaro-Almagro, F., Sotiropoulos, S. N., ... Andersson, J. L. R. (2019). Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *NeuroImage*, 184(January), 801–812. <https://doi.org/10.1016/j.neuroimage.2018.09.073>
- Brunet, D., Vrscaj, E. R., & Wang, Z. (2012). On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, 21(4), 1488–1499. <https://doi.org/10.1109/TIP.2011.2173206>
- Charrier, C., Knoblauch, K., Maloney, L. T., Bovik, A. C., & Moorthy, A. K. (2012). Optimizing Multiscale SSIM for Compression via MLDS. *IEEE Transactions on Image Processing*, 21(12), 4682–4694. <https://doi.org/10.1109/TIP.2012.2210723>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining—KDD '16* (pp. 785–794). San Francisco, CA: ACM Press. <https://doi.org/10.1145/2939672.2939785>
- Chow, L. S., & Paramesran, R. (2016). Review of medical image quality assessment. *Biomedical Signal Processing and Control*, 27(May), 145–154. <https://doi.org/10.1016/j.bspc.2016.02.006>
- Daszykowski, M., Walczak, B., & Massart, D. L. (2002). Looking for natural patterns in analytical data. 2. Tracing local density with OPTICS. *Journal of Chemical Information and Computer Sciences*, 42(3), 500–507. <https://doi.org/10.1021/ci010384s>
- David, A. & Vassilvitskii, S. (2007). K-Means++: The advantages of careful seeding. 1027–1035.
- David, S., Mesri, H. Y., Viergever, M. A., & Leemans, A. (2019). Statistical significance in DTI group analyses: How the choice of the estimator can inflate effect sizes. *BioRxiv* 755140. <https://doi.org/10.1101/755140>
- de Lange, A.-M. G., Barth, C., Kaufmann, T., Maximov, I. I., van der Meer, D., Agartz, I., & Westlye, L. T. (2020). Women's brain aging: Effects of sex-hormone exposure, pregnancies, and genetic risk for Alzheimer's disease. *Human Brain Mapping*, 41(18), 5141–5150. <https://doi.org/10.1002/hbm.25180>
- Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLOS ONE*, 10(4), e0121945. <https://doi.org/10.1371/journal.pone.0121945>
- Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987). Principles of nuclear magnetic resonance in one and two dimensions. In *The international series of monographs on chemistry* 14. Oxford [Oxfordshire]: New York: Clarendon Press; Oxford University Press.
- Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)*. AAAI Press, 226–231.

- Fieremans, E., Jensen, J. H., & Helpert, J. A. (2011). White matter characterization with diffusional kurtosis imaging. *NeuroImage*, 58(1), 177–188. <https://doi.org/10.1016/j.neuroimage.2011.06.006>
- Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., ... Shinohara, R. T. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, 161, 149–170. <https://doi.org/10.1016/j.neuroimage.2017.08.047>
- Graham, M. S., Drobnyak, I., & Zhang, H. (2018). A supervised learning approach for diffusion MRI quality control with minimal training data. *NeuroImage*, 178, 668–676. <https://doi.org/10.1016/j.neuroimage.2018.05.077>
- Grinberg, F., Maximov, I. I., Farrher, E., Neuner, I., Amort, L., Thönneßen, H., ... Jon Shah, N. (2017). Diffusion kurtosis metrics as biomarkers of microstructural development: A comparative study of a Group of Children and a Group of Adults. *NeuroImage*, 144(January), 12–22. <https://doi.org/10.1016/j.neuroimage.2016.08.033>
- Haddad, S. M. H., Scott, C. J. M., Ozzoude, M., Holmes, M. F., Arnott, S. R., Nanayakkara, N. D., ... Bartha, R. (2019). Comparison of quality control methods for automated diffusion tensor imaging analysis pipelines. *PLoS One*, 14, e0226715.
- Jelescu, I. O., Veraart, J., Fieremans, E., & Novikov, D. S. (2016). Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue: Degeneracy in model parameter estimation of diffusion in neural tissue. *NMR in Biomedicine*, 29(1), 33–47. <https://doi.org/10.1002/nbm.3450>
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, 62(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- Jensen, J. H., Helpert, J. A., Ramani, A., Lu, H., & Kaczynski, K. (2005). Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine*, 53(6), 1432–1440. <https://doi.org/10.1002/mrm.20508>
- Jones, D. K. (Ed.). (2010). *Diffusion MRI: Theory, methods, and application*. Oxford, New York: Oxford University Press.
- Kaufmann, T., van der Meer, D., Doan, N. T., Schwarz, E., Lund, M. J., Agartz, I., ... Westlye, L. T. (2019). Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience*, 22(10), 1617–1623. <https://doi.org/10.1038/s41593-019-0471-7>
- Kellner, E., Dhital, B., Kiselev, V. G., & Reiser, M. (2016). Gibbs-ringing artifact removal based on local subvoxel-shifts: Gibbs-ringing artifact removal. *Magnetic Resonance in Medicine*, 76(5), 1574–1581. <https://doi.org/10.1002/mrm.26054>
- Koch, A., Zhukov, A., Stöcker, T., Groeschel, S., & Schultz, T. (2019). SHORE-based detection and imputation of dropout in diffusion MRI. *Magnetic Resonance in Medicine*, 82(6), 2286–2298. <https://doi.org/10.1002/mrm.27893>
- Kochunov, P., Dickie, E. W., Viviano, J. D., Turner, J., Kingsley, P. B., Jahanshad, N., ... Voineskos, A. N. (2018). Integration of routine QA data into mega-analysis may improve quality and sensitivity of multi-site diffusion tensor imaging studies. *Human Brain Mapping*, 39(2), 1015–1023. <https://doi.org/10.1002/hbm.23900>
- Kochunov, P., Jahanshad, N., Marcus, D., Winkler, A., Sprooten, E., Nichols, T. E., ... van Essen, D. C. (2015). Heritability of fractional anisotropy in human white matter: A comparison of human connectome project and ENIGMA-DTI data. *NeuroImage*, 111(May), 300–311. <https://doi.org/10.1016/j.neuroimage.2015.02.050>
- Kolmogoroff, A. N. (1933). Sulla Determinazione Empirica Di Una Legge Di Distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4, 81–91.
- Lebel, C., Gee, M., Camicioli, R., Wieler, M., Martin, W., & Beaulieu, C. (2012). Diffusion tensor imaging of white matter tract evolution over the lifespan. *NeuroImage*, 60(1), 340–352. <https://doi.org/10.1016/j.neuroimage.2011.11.094>
- Li, C., & Bovik, A. (2009). Three-component weighted structural similarity index. *Proceedings of SPIE-IS&T Electronic Imaging*, 7242. <https://doi.org/10.1117/12.811821>
- Maximov, I. I., Alnæs, D., & Westlye, L. T. (2019). Towards an optimised processing pipeline for diffusion magnetic resonance imaging data: Effects of artefact corrections on diffusion metrics and their age associations in UKbiobank. *Human Brain Mapping*, 40(14), 4146–4162. <https://doi.org/10.1002/hbm.24691>
- Maximov, I. I., Thönneßen, H., Konrad, K., Amort, L., Neuner, I., & Jon Shah, N. (2015). Statistical instability of TBSS analysis based on DTI fitting algorithm: TBSS analysis. *Journal of Neuroimaging*, 25(6), 883–891. <https://doi.org/10.1111/jon.12215>
- Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Junqian, X., et al. (2016). Multimodal population brain imaging in the UKbiobank prospective epidemiological study. *Nature Neuroscience*, 19(11), 1523–1536. <https://doi.org/10.1038/nn.4393>
- Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., ... Rathi, Y. (2018). Multi-site harmonization of diffusion MRI data in a registration framework. *Brain Imaging and Behavior*, 12(1), 284–295. <https://doi.org/10.1007/s11682-016-9670-y>
- Muckley, M. J., Ades-Aron, B., Papaioannou, A., Lemberskiy, G., Solomon, E., Lui, Y. W., ... Knoll, F. (2021). Training a neural network for Gibbs and noise removal in diffusion MRI. *Magnetic Resonance in Medicine*, 85, 413–428. <https://doi.org/10.1002/mrm.28395>
- Novikov, D. S., Kiselev, V. G., & Jespersen, S. N. (2018). On modeling. *Magnetic Resonance in Medicine*, 79(6), 3172–3193. <https://doi.org/10.1002/mrm.27101>
- Novikov, D. S., Veraart, J., Jelescu, I. O., & Fieremans, E. (2018). Rotationally-invariant mapping of scalar and orientational metrics of neuronal microstructure with diffusion MRI. *NeuroImage*, 174, 518–538. <https://doi.org/10.1016/j.neuroimage.2018.03.006>
- Perrone, D., Aelterman, J., Pizurica, A., Jeurissen, B., Phillips, W., & Leemans, A. (2016). The effect of Gibbs ringing artifacts on measures derived from diffusion MRI. *NeuroImage*, 120, 441–455. <https://doi.org/10.1016/j.neuroimage.2015.06.068>
- Renieblas, G. P., Nogués, A. T., González, A. M., Gómez-Leon, N., & del Castillo, E. G. (2017). Structural similarity index family for image quality assessment in radiological images. *Journal of Medical Imaging*, 4(3), 035501. <https://doi.org/10.1117/1.JMI.4.3.035501>
- Richard, G., Kolskår, K., Sanders, A.-M., Kaufmann, T., Petersen, A., Doan, N. T., et al. (2018). Assessing distinct patterns of cognitive aging using tissue-specific brain age prediction based on diffusion tensor imaging and brain morphometry. *PeerJ*, 6(November), e5908. <https://doi.org/10.7717/peerj.5908>
- Rudrapatna, U., Parker, G. D., Roberts, J., & Jones, D. K. (2020). A comparative study of gradient nonlinearity correction strategies for processing diffusion data obtained with ultra-strong gradient MRI scanners. *Magnetic Resonance in Medicine*, 85, 1104–1113. <https://doi.org/10.1002/mrm.28464>
- Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 19(2), 279–281. <https://doi.org/10.1214/aoms/1177730256>
- Smith, S. M., Johansen-Berg, H., Jenkinson, M., Rueckert, D., Nichols, T. E., Miller, K. L., ... Behrens, T. E. J. (2007). Acquisition and Voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. *Nature Protocols*, 2(3), 499–503. <https://doi.org/10.1038/nprot.2007.45>
- Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E., & Miller, K. L. (2019). Estimation of brain Age Delta from brain imaging. *NeuroImage*, 200(October), 528–539. <https://doi.org/10.1016/j.neuroimage.2019.06.017>
- Tabesh, A., Jensen, J. H., Ardekani, B. A., & Helpert, J. A. (2010). Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. *Magnetic Resonance in Medicine*, 65, 823–836. <https://doi.org/10.1002/mrm.22655>

- Tournier, J., Smith, R. E., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., ... Connelly, A. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage*, 202.
- Tønnesen, S., Kaufmann, T., Doan, N. T., Alnæs, D., Córdova-Palamera, A., van der Meer, D., et al. (2018). White matter aberrations and age-related trajectories in patients with schizophrenia and bipolar disorder revealed by diffusion tensor imaging. *Scientific Reports*, 8(1), 14129. <https://doi.org/10.1038/s41598-018-32355-9>
- Veraart, J., Fieremans, E., Jelescu, I. O., Knoll, F., & Novikov, D. S. (2016). Gibbs ringing in diffusion MRI. *Magnetic Resonance in Medicine*, 76, 301–314. <https://doi.org/10.1002/mrm.25866>
- Veraart, J., Novikov, D. S., Christiaens, D., Ades-aron, B., Sijbers, J., & Fieremans, E. (2016). Denoising of diffusion MRI using random matrix theory. *NeuroImage*, 142, 394–406. <https://doi.org/10.1016/j.neuroimage.2016.08.016>
- Veraart, J., Novikov, D. S., & Fieremans, E. (2018). TE dependent diffusion imaging (TEdDI) distinguishes between compartmental T 2 relaxation times. *NeuroImage*, 182, 360–369. <https://doi.org/10.1016/j.neuroimage.2017.09.030>
- Veraart, J., Sijbers, J., Sunaert, S., Leemans, A., & Jeurissen, B. (2013). Weighted linear least squares estimation of diffusion MRI parameters: Strengths, limitations, and pitfalls. *NeuroImage*, 81, 335–346. <https://doi.org/10.1016/j.neuroimage.2013.05.028>
- Veraart, J., Van Hecke, W., & Sijbers, J. (2011). Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a Rician noise model. *Magnetic Resonance in Medicine*, 66, 678–686. <https://doi.org/10.1002/mrm.22835>
- Vinding, M. S., Brenner, D., Tse, D. H. Y., Vellmer, S., Vosegaard, T., Suter, D., ... Maximov, I. I. (2017). Application of the limited-memory quasi-Newton algorithm for multi-dimensional, large Flip-angle RF pulses at 7T. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 30(1), 29–39. <https://doi.org/10.1007/s10334-016-0580-1>
- Vos, S. B., Tax, C. M. W., Luijten, P. R., Ourselin, S., Leemans, A., & Froeling, M. (2016). The importance of correcting for signal drift in diffusion MRI. *Magnetic Resonance in Medicine*, 77, 285–299. <https://doi.org/10.1002/mrm.26124>
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. <https://doi.org/10.1109/TIP.2003.819861>
- Westlye, L. T., Walhovd, K. B., Dale, A. M., Bjørnerud, A., Due-Tønnessen, P., Engvig, A., ... Fjell, A. M. (2010). Life-span changes of the human brain white matter: Diffusion tensor imaging (DTI) and Volumetry. *Cerebral Cortex*, 20(9), 2055–2068. <https://doi.org/10.1093/cercor/bhp280>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Maximov I, van der Meer D, de Lange A-MG, et al. Fast quality control method for derived diffusion Metrics (YTRIUM) in big data analysis: U.K. Biobank 18,608 example. *Hum Brain Mapp*. 2021;42: 3141–3155. <https://doi.org/10.1002/hbm.25424>