



SAHLGRENSKA ACADEMY

DEFORMABLE IMAGE REGISTRATION OF SPINE MAGNETIC RESONANCE IMAGES WITH ELASTIX FOR DETAILED CHARACTERIZATION OF DISC LOADING BEHAVIORS

M. Sc. Thesis

Zainab Sirat

| | |
|------------------------|--|
| Essay/Thesis: | 30 hp |
| Program and/or course: | Medical Physics |
| Level: | Second Cycle |
| Semester/year: | Spring 2020 |
| Supervisor: | Kerstin Lagerstrand, Fredrik Nordström |
| Examiner: | Magnus Båth |

Abstract

Essay/Thesis: 30 hp
Program and/or course: Medical Physics
Level: Second Cycle
Semester/year: Spring 2020
Supervisor: Kerstin Lagerstrand, Fredrik Nordström
Examiner: Magnus Båth
Keyword: Registration, MRI, Spine

Purpose: This work aimed to develop an optimized strategy for registration of spine magnetic resonance (MR) images acquired with and without spinal loading, as well as evaluate the quality of the registration strategy with commonly used quality assessment methods.

Method: Previously collected MR images of the lumbar spine of 37 individuals were re-analyzed. Multi-echo T2-mapping had been obtained from the subjects in both unloaded and axial-loaded condition. The signal intensity-based registration software Elastix were selected for the registrations between the two image data sets. To facilitate automatic and advanced image analysis, the MICE Toolkit was used as a graphical user interface for Elastix. The registration was done in two steps using the short echo-time raw data images: (1) rigid registration and (2) deformable registration (*i.e.* non-rigid) with the first registration as input. For comparison, registration was also performed using the long echo-time raw data images. To reduce the calculation time and possibly improve the quality of the image registration, the image was limited by a binary image mask that was applied over the spinal column. To optimize the registration, the parameters FinalGridSpacingInVoxels and BSplineInterpolationOrder were varied. The registration quality was evaluated with Dice similarity coefficient (DSC) and Jaccard coefficient, where a semi-automatic software based on region-growing was used for the delineation of the intervertebral discs (IVDs). The Jacobian determinant was also calculated to ensure that the deformation was realistic with non-zero positive values. The intra- and inter-observer reliability between readers were also determined.

Result: The DSC (0.845 ± 0.059) and the Jaccard coefficient (0.735 ± 0.084) were high for all individuals. The mean of the Jacobian determinant was close to one (1.035 ± 0.043). Analysis based on the whole image stack, also showed high DSC (0.845 ± 0.057) and the Jaccard coefficient (0.736 ± 0.081). Significant difference was neither observed between registration with short and long TE images ($p = 0.205$), nor between registration with or without mask ($p = 0.247$). A slightly higher accuracy was obtained with less final B-spline grid spacing. The inter- and intra-observer agreement were excellent ($ICC = 0.99$)

Conclusion: Deformable image registration of spine MR images with the optimized registration strategy seems feasible and useful for detailed characterization of disc loading behaviours. The registration using the strategy showed high quality evaluated with DSC and Jaccard coefficient. The evaluation of the Jacobian determinant indicated a preserved topology of the deformed image.

Table of content

| | |
|---|----|
| 1 Background | 1 |
| 2 Material and methods | 3 |
| 2.1 Study population/Cohort | 3 |
| 2.2 Magnetic resonance imaging..... | 3 |
| 2.3 The strategy for registration of images with unloaded and loaded spine | 4 |
| 2.4 Segmentation of images with unloaded and loaded spine..... | 4 |
| 2.4.1 Segmentation of three innermost slices for all 37 individuals..... | 4 |
| 2.4.2 Difference in registration between short and long TE..... | 5 |
| 2.4.3 Segmentation of nine slices for six individuals..... | 5 |
| 2.5 Image registration process | 6 |
| 2.6 Reduction of the registration region by masking for improved image quality | 6 |
| 2.7 Optimization of non-rigid registration step | 7 |
| 2.8 Repeatability | 7 |
| 3 Result | 8 |
| 3.1.1 Difference in registration between the whole image stack and the three innermost slices | 8 |
| 3.1.2 Difference in registration between long and short echo time images..... | 8 |
| 3.2 Influence of mask on the registration quality | 9 |
| 3.3 Optimization of non-rigid registration step | 11 |
| 3.4 Repeatability..... | 11 |
| 4 Discussion | 11 |
| 5 Conclusion | 13 |
| Reference list | 15 |
| Appendix | 17 |

1 Background

Low back pain (LBP) arises intrinsically from the spine, intervertebral discs (IVDs) or surrounding soft tissue, which could be caused by pathological degenerative changes in the lumbar spine such as inflammatory, genetic, infectious, fissures in the annulus fibrosus (Figure 1) or fatigue injury of the vertebral endplates. The IVD undergoes degenerative morphological and cellular changes with age [1] and the degeneration affects the IVD's functionality with reduced capacity to resist load. The degeneration of the IVD may cause changes in the stress pattern of the annulus fibrosus and the vertebral endplates that may lead to matrix damage exceeding the body's ability to heal. Most degeneration changes are asymptomatic. However, IVD degeneration is shown to be associated with LBP [2].

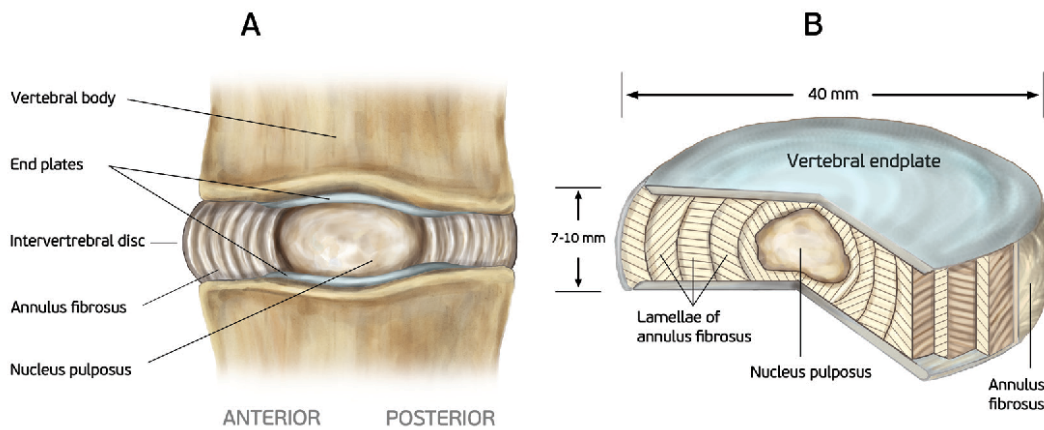


Figure 1. Location and structure of the nucleus pulposus, annulus fibrosus inside the IVD and endplates on (A) a sagittal section of the spine and (B) on a cutout portion of the disc.
(Tomaszewski et al. 2015. The biology behind the human intervertebral disc and its endplates, figure 1, Folia Morphol (Warsz), doi:10.5603/FM.2015.0026)

At current, LBP patients are routinely investigated in the supine position using conventional magnetic resonance imaging (MRI) sequences with the spine unloaded. However, conventional MRI is deficient for differentiating a painful degenerated IVD from a degenerated but asymptomatic IVD [3, 4]. Functional MRI is believed to improve the diagnostic accuracy of LBP and provide increased knowledge about the disease. The complex nature of the IVD can, for example, be studied *in vivo* during the influence of axial load, which is theoretically appealing since most patients with LBP experience more symptoms during an upright position when loading the spine [5]. Also, the quantitative MRI T2-mapping technique, applied with and without loading of the spine, has been shown to characterize different alteration in the dorsal regions of the IVDs between LBP-patients and asymptomatic individuals from a control group [2].

T2-mapping images (T2-maps) are calculated from multiple image sets, representing different echo times (TE). The signal intensity of each pixel in the multiple TE images is fitted to a model of exponential signal decay to obtain the T2-time [6].

$$S_i = M_0 \exp\left(\frac{-TE_i}{T_2\text{-time}}\right), (1)$$

where S_i is the signal intensity of the pixel, and M_0 is the equilibrium magnetization weighted with the receiver coil sensitivity.

The T2-time has been shown to depict the structural integrity of the disc matrix [7], displaying a stronger correlation with collagen structure than with changes in the cartilage matrix of the IVD. Hence, the regional variation of the T2-value over the IVD acquired with no loading and loading of the spine may reveal dynamical information likely due to changes in intradiscal water disposition and tissue anisotropy [2, 8, 9].

To be able to characterize the regional variation of the loading effect ($T2\text{-value}_{\text{with load}} - T2\text{-value}_{\text{without load}}$) in more detail and, thus, enable analysis on a voxel-by-voxel basis, deformable registration of images acquired with and without loading needs to be performed. Since such detailed characterization is dependent on the quality of the registration, high performance software, as well as optimized strategies for the registration are needed.

In the literature, the open source software Elastix is well described and thoroughly validated and has been pointed out as a high-performance registration tool [10, 11]. Elastix was developed to facilitate research on medical images [12]. The software is built upon the widely used Insight Toolkit (ITK) and includes several optimization methods, interpolators, transformation models and cost functions. The main idea of this intensity-based registration software is to search iteratively for the geometric transformation that, when applied to the moving image, optimizes a similarity measure (cost function) that is related to voxel intensity and is computed in the overlapping regions of the input image [13]. Hence, Elastix should be suitable for deformable image registration (DIR) of spine images acquired with and without axial loading of the spine. To our knowledge, however, no previous work has demonstrated its feasibility.

The performance of the registration can be presented in terms of different quantitative measures. Visual inspection of the registered images gives the observer a rough picture of the quality of the registration, but only qualitatively. To ensure that the registration of the images is of high quality, quantitative assessments should be performed, preferably with methods that allow for direct validation of the registration. Registrations are commonly evaluated using the Dice similarity coefficient (DSC) and the Jaccard coefficient. DSC takes the ratio of the overlapping volume over the total volume of both regions and Jaccard coefficient uses the union volume as the denominator for both regions, and also manages the deficiency of target overlap [14]. The DSC and Jaccard coefficient rely on image segmentations of the spinal structures and measure the similarity between finite sample sets, *e.g.* the overlap between segmented IVDs. The Jacobian determinant, which measures the relative volume change of the deformed voxel. A value of one corresponds to zero contraction or expansion of a point in the deformation [14], and a negative value of the Jacobian determinant means that the orientation is reversed by the transformation, *i.e.* unrealistic deformation.

This work aimed to 1) develop an optimized registration strategy for future detailed characterization of the loading effect on the IVD and 2) evaluate the quality of the registration strategy with commonly used quality assessment methods.

2 Material and methods

2.1 Study population/Cohort

This study was a secondary-analysis of previously collected MRI aiming to characterize different alternation in the IVDs during loading, between LBP-patients and asymptomatic individuals (see [2] for further information).

Twenty-six LBP patients (25-69 years, mean 38 years, 11 males) and 12 controls (25-59 years, mean 38 years, 7 males) were included. All patients suffered from chronic LBP with at least six months of continuous pain of non-specific character. Individuals with bulging, herniation or degenerated IVD disease where not included in the study.

The study has been approved by the regional ethics review board in Gothenburg (Dnr 888-14).

2.2 Magnetic resonance imaging

The lumbar spine (L1 to S1) of all individuals were examined on a 1.5 Tesla MRI scanner (Siemens Magnetom Aera, Erlangen, Germany) in both unloaded and axial-loaded condition. All subjects were examined with sagittal T1-weighted (T1W) and T2-weighted (T2W) imaging, and additionally with quantitative T2-mapping. The T2 mapping was performed using a multi-echo spin-echo sequence with eight different TEs varying from 11.1 ms to 88.8 ms. Fundamental parameters for all sequences are described in Table 1.

Table 1. MRI scanning parameters for T1W, T2W and T2-map.

| | T1W (TSE) | T2W (TSE) | T2W (TSE) | T2-map (MESE) |
|---------------------------------------|------------------|------------------|------------------|--|
| Imaging plane | Sagittal | Sagittal | Axial | Sagittal |
| Repetition time | 480 | 3500 | 4862 | 1400 |
| Echo time (ms) | 9 | 95 | 97 | 11.1, 22.2, 33.3, 44.4, 55.5, 66.6, 77.7, 88.8 |
| Echo train length | 3 | 14 | 17 | 8 |
| Slice thickness (mm) | 3.5 | 3.5 | 3.5 | 3.5 |
| Slice gap (mm) | 0.7 | 0.7 | 0.4 | 0.7 |
| Number of signals averaged | 2 | 1 | 2 | 1 |
| Pixel bandwidth (Hz) | 235 | 180 | 195 | 220 |
| Flip angle (degree) | 150 | 150 | 150 | 180 |
| Acquisition matrix | 320×224 | 384×288 | 320×256 | 256×256 |
| Reconstruction matrix | 320×320 | 300×384 | 320×320 | 256×256 |
| Field of view (mm²) | 300×300 | 300×300 | 200×200 | 220×220 |

The axial loading of the spine was performed using a compression device (DynaWell Diagnostics Inc, Las Vegas, NV, USA) pursued to simulate loadbearing during supine imaging, illustrated in Figure 2. During compression the individual laid supine with extended hips and knees with applied load corresponding to 50% of the body weight to simulate the loading forces the spine is exposed to in an upright position [2].



Figure 2. Illustration of the compression device for axial loading of the spine.

(Healthcare in Europe, <https://healthcare-in-europe.com/en/news/ct-mri-compatible-compression-harness.html>
Image courtesy of DynaWell Diagnostics. Accessed July 27th 2020)

2.3 Strategy for registration of images with unloaded and loaded spine

An optimized strategy for registration of images without loading and with loading of the spine was developed to enable detailed characterization of the loading effect in the IVDs voxel-by-voxel.

For that purpose, a subset of the multi-echo raw data images that were acquired for calculation of the T2-maps were used. These images were acquired in the same image space but at different echo times. Transformations from registrations of images with short echo time can, therefore, be applied to the long echo time images or to the T2-map directly. As a consequence, the images with the highest and best suitable contrast can be chosen for the registration.

Thus, to enable registration of the deformation of the IVD while avoiding the loading effect on signal intensities in the late echo images, the short-echo images ($T2_{raw,11.1}$) that display T1-weighted contrast were chosen for the registration.

The signal intensity-based registration software Elastix (Software 5.0.0, University Medical Center Utrecht, The Netherlands) [15] was selected for the registrations. It is a well described and thoroughly validated open source software [16], which has been proven feasible for deformable registration of MR images. The MICE Toolkit (Toolkit 1.1.0, NONPI Medical AB, Umeå, Sweden) was used as a graphical user interface for Elastix facilitating automatic advanced image analysis. The default settings in MICE for Elastix registration were used as listed in Appendix. If needed, these settings were planned to be adjusted in a sub-analysis (see section 2.7).

2.4 Segmentation of images with unloaded and loaded spine

2.4.1 Segmentation of three innermost slices for all 37 individuals

To evaluate the registration quality using DSC and Jaccard coefficient, each IVD was segmented semi-automatically with an in-house developed software based on region-growing (see [17] for further details). The shortest echo raw data images ($T2_{raw,11.1}$) with highest signal to noise ratio and lowest signal variation over the IVD were used for the segmentation.

Moreover, the short-echo raw data images were chosen for the segmentation as they had a contrast that was close to T1W and as such displayed the IVDs with high contrast edges and

flatter contrast of the inner structure, improving the segmentation of the IVDs. Only the three innermost slices of the sagittal slices were selected to reduce the time-consuming segmentation process. The selection of the sagittal slices was enabled using the RadiAnt DICOM Viewer 5.5.1 software (Medixant, Poznan, Poland), where the position of the sagittal slices in relation to the center of the IVD were displayed on conventional axial T1W-images.

The segmentation was performed combining the region-growing algorithm with manual adjustment. The outer contour of each IVD in the lumbar spine region (L1 to L5) was outlined on the three mid-sagittal $T2_{\text{raw},11.1}$ -images. The procedure was performed both for images with and without loading of the spine. Care was taken to exclude tissues not associated with the IVD (Figure 3). The segmented “regions of interest” (ROIs) were discretized to binary masks and these masks were then exported in Neuroimaging Informatics Technology Initiative (NifTI)-format and used for calculation of the DSC and the Jaccard coefficient both for the short-echo and long-echo raw data images. All DSC values and Jaccard coefficients are presented as mean \pm standard deviation [min-max], also the computation time was presented. The ROIs were also saved for future evaluation of the loading effecting using registration based T2-mapping.

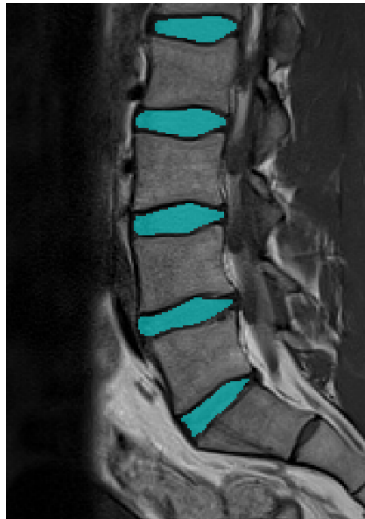


Figure 3. A case point of segmented IVDs used for the registration.

2.4.2 Difference in registration between short and long TE

For comparison, not only registrations of images with T2-weighted contrast ($T2_{\text{raw},11.1}$) but also registrations of images with strong T2-weighted contrast, *i.e.* the long-echo raw data images ($T2_{\text{raw},88.8}$) and hence high visibility of water and matrix changes were performed. To determine whether there is any difference in the quality of the registration between long and short echo images, t-test with 95% confidence interval was performed with $p < 0.05$ representing significant difference.

2.4.3 Segmentation of nine innermost slices for six individuals

To evaluate the impact of slice position for the registration quality, a subsample of the study cohort was chosen randomly. In these six individuals, all slices *i.e.* 9 slices of the image stack were used for the IVD segmentation. The segmentation of the IVDs were performed in the same way as described in section 2.4.1. Difference in the DSC and Jaccard coefficient was then evaluated statistically using a paired t-test, with $p < 0.05$ representing significant difference. Test for equivalence was also performed using the parametric two one-sided-test (TOST) with 95% confidence interval.

2.5 Image registration process

The registration was done in two steps, first a rigid registration was performed and the output *i.e.* the rigidly registered image was input to the next step. The rigid transformation was performed using EulerTransform with other default parameters such as e.g. NumberOfResolutions 4, MaximumNumberOfIterations 250 and FinalBSplineInterpolationOrder 3. The next step was a non-rigid and deformable registration, which returned the deformed image as well as Elastix transform parameters as output. The non-rigid transformation was performed using BSplineTransform, FinalGridSpacingInVoxels 16, NumberOfResolutions 4, MaximumNumberOfIterations 500 and BSplineInterpolationOrder 1 (see Appendix for more details). To evaluate the impact of the image contrast, the same procedure was performed also with the $T2_{\text{raw},88.8}$ -images.

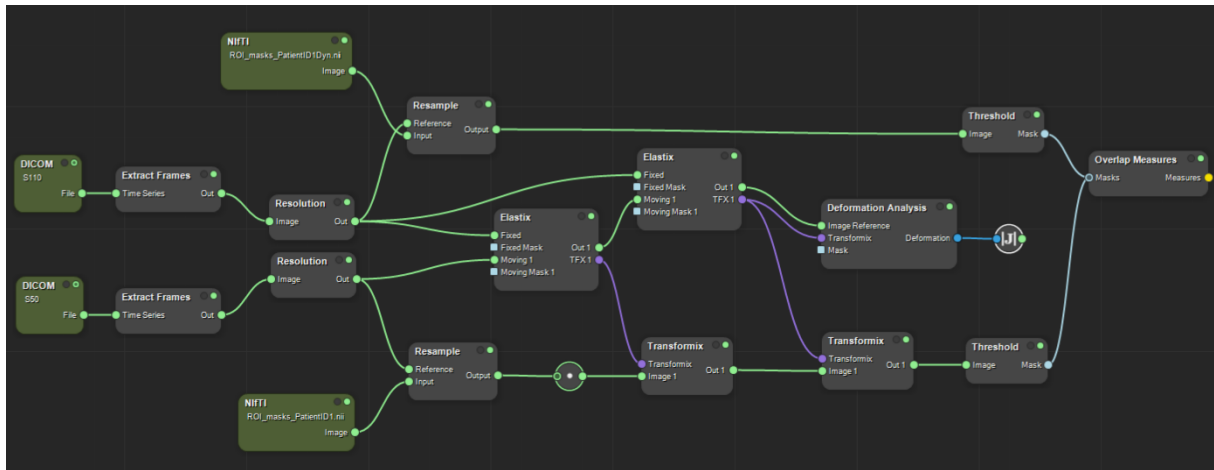


Figure 4. The code for registration with overlap measures for the DSC and Jaccard coefficient calculation and deformation analysis for the Jacobian determinant calculation.

DICOM folders with the short-echo raw data images with 13 slices in every stack for both the loaded and unloaded condition of the spine were imported into MICE (Figure 4). After extracting 3D-images of relevant echo time ($T2_{\text{raw},11.1}$), image matrix resolution was increased to 1 mm isotropic voxels using linear interpolation. The images of the loaded spine were set to fixed and the unloaded were set to moving in the registration steps. Each registration step generated two outputs, an Elastix transform parameter set and an image, *i.e.* of the loaded spine to which the transformation has been applied. The output image of the rigid registration step was used as the input to the DIR step. Visual assessment, *i.e.* comparing the output image of the DIR with the image of the unloaded spine, was undertaken.

In order to quantify the quality of the registration, the transform parameters were used to deform the mask of the loaded spine in the same manner as the image. The output image was converted to a binary mask using thresholding and the overlap accuracy was calculated, in form of DSC and Jaccard coefficient, where a value of 1 means perfect overlap. For comparison, both DSC and Jaccard coefficient were chosen to be presented for other studies to compare their results depending on quality assessment method. Finally, Jacobian determinants were calculated from the outputs of the DIR.

2.6 Reduction of the registration region by masking for improved image quality

To reduce the calculation time and possibly improve the quality of the image registration, a rectangular mask was applied over the spinal column (Figure 5). Only the masked region was used in the registration and compared with registration using the corresponding unmasked

image. The quality of the registration was evaluated visually by comparing the registered and fixed images. Also, the computation time was presented.

Additionally, the DSC and the Jaccard coefficient between the registered and the fixed images were compared. A paired t-test with one tail was performed to determine significant difference at $p > 0.05$. Test for similarity was also performed using TOST.

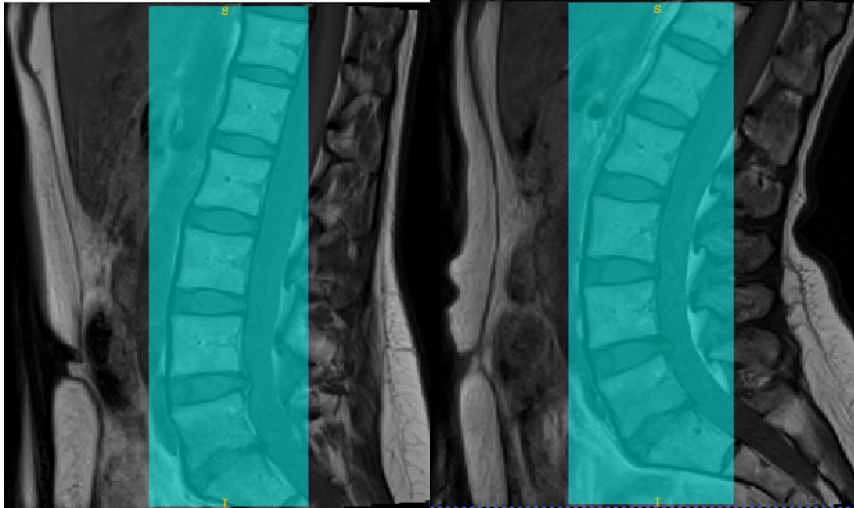


Figure 5. Illustration of the masks applied upon the loaded and unloaded spine before registration.

2.7 Optimization of the non-rigid registration step

The B-spline transformation is defined by a uniform grid of control points. This grid is defined by the spacing between the grid nodes which defines how dense the grid is, or what the locality is of the transformation that is being modelled. A different grid spacing can be defined for each resolution level, *i.e.* the idea is that subsequently match smaller structures, up to final precision. A high value of the final B-spline grid spacing entails that small structures cannot be matched. A very low value of the grid spacing will on the other hand make small structures match but also allows the transformation to have more freedom, which could result in irregular transformations (unrealistic deformations) in homogeneous parts of the image [18]. By default, the grid spacing is halved after every resolution such that the final grid spacing is obtained in the last resolution level. The parameter `FinalGridSpacingInVoxels` was varied from 4 to 32 and was reported in quality parameters DSC and Jaccard coefficient, as well as computation time.

The parameter `BSplineInterpolationOrder` is the order of B-spline interpolation used for applying the first deformation, and was changed from first-order to third, to evaluate how much better the quality of the registration becomes. A first-order B-spline interpolation, *i.e.* linear interpolation gives satisfactory results with a good trade-off between quality and speed of the registration [18]. The higher the order, the better the quality but also the more computation time is required.

2.8 Repeatability

The intra- and inter-observer reliability between readers have previously been determined for MRI images of the spine with a high degree of consensus (ranging between 0,79 to 0,99 depending on the ROI evaluated) [2]. To verify this for the actual reader (Z.S.), the segmentation was repeated on a subset of individuals (6 individual of the study cohort), both

for unloaded and loaded conditions. The intraclass correlation coefficient (ICC) was calculated with 95% confidence intervals. Values < 0.5 was regarded as poor, 0.5 to 0.75 as moderate, 0.75 to 0.9 as good and > 0.90 as excellent [19].

3 Result

3.1.1 Difference in registration between the whole image stack and the three innermost slices
Based on the three image slices, the DSC (0.845 ± 0.059 [0.682-0.907]) and the Jaccard coefficient (0.735 ± 0.084 [0.517-0.830]) were high for all individuals (Figure 6). The mean of the Jacobian (1.035 ± 0.043 [0.515-1.773]) was close to one and negative values were not observed.

Analysis based on the whole image stack, also showed high DSC and the Jaccard coefficient (0.827 ± 0.032 [0.778-0.868] and 0.706 ± 0.078 [0.637-0.766], respectively). The mean of the Jacobian determinant (1.026 ± 0.043 [0.966-0.1084]) was also here close to one. No significant difference was observed, with a p-value of 0.294. The TOST test showed equivalence at 95% confidence interval between registration with 3 slices and using the whole stack.

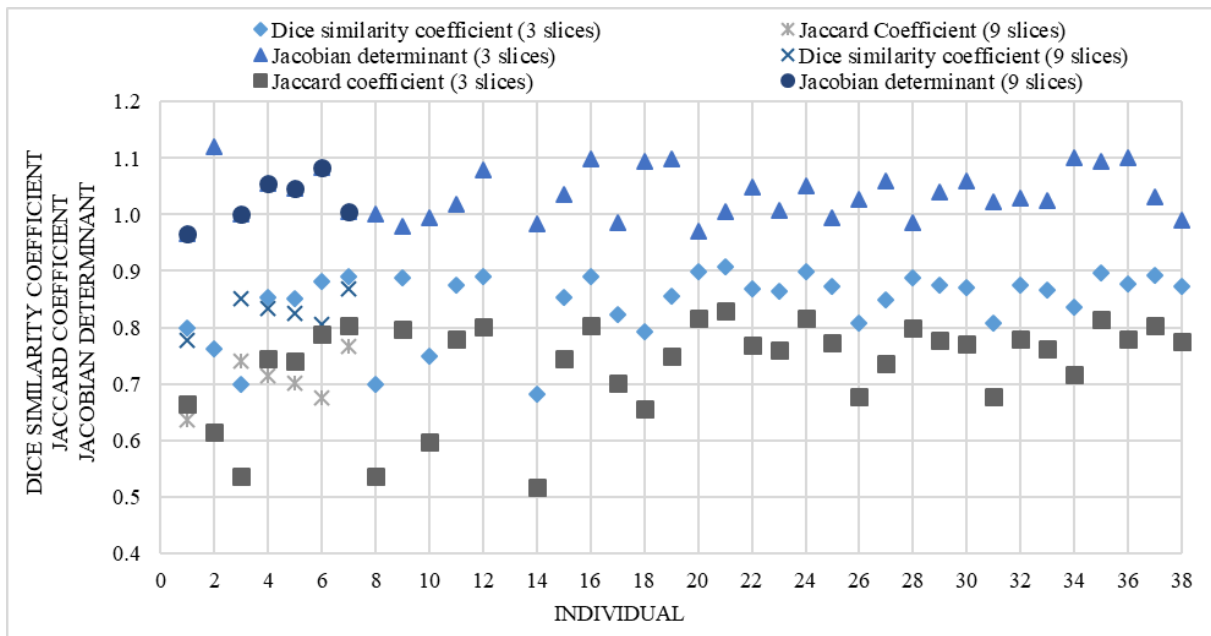


Figure 6. The DSC, Jaccard coefficient, and mean of Jacobian determinant for the included individuals using three slices, as well as DSC and Jaccard coefficient for 6 using the whole stack, *i.e.* 9 slices.

3.1.2 Difference in registration between long and short echo time images

For registration with $T_{2,raw,88.8}$ -images the DSC (0.847 ± 0.054 [0.707-0.902]) and the Jaccard coefficient (0.738 ± 0.077 [0.547-0.822]) were high for all individuals.

The DSC and Jaccard coefficient for registration with $T_{2,raw,11.1}$ -images and $T_{2,raw,88.8}$ -images are presented as a boxplot in Figure 7. The mean, median, first and third quartile for DSC were 0.845, 0.869, 0.824, 0.887 and 0.847, 0.867, 0.831, 0.880, respectively. The corresponding values for the Jaccard coefficient were 0.735, 0.769, 0.700, 0.798 and 0.738, 0.765, 0.711, 0.785, respectively.

The p-value for the paired t-test was calculated to 0.205, *i.e.* no significant difference was observed. The TOST test showed equivalence at 95% confidence interval.

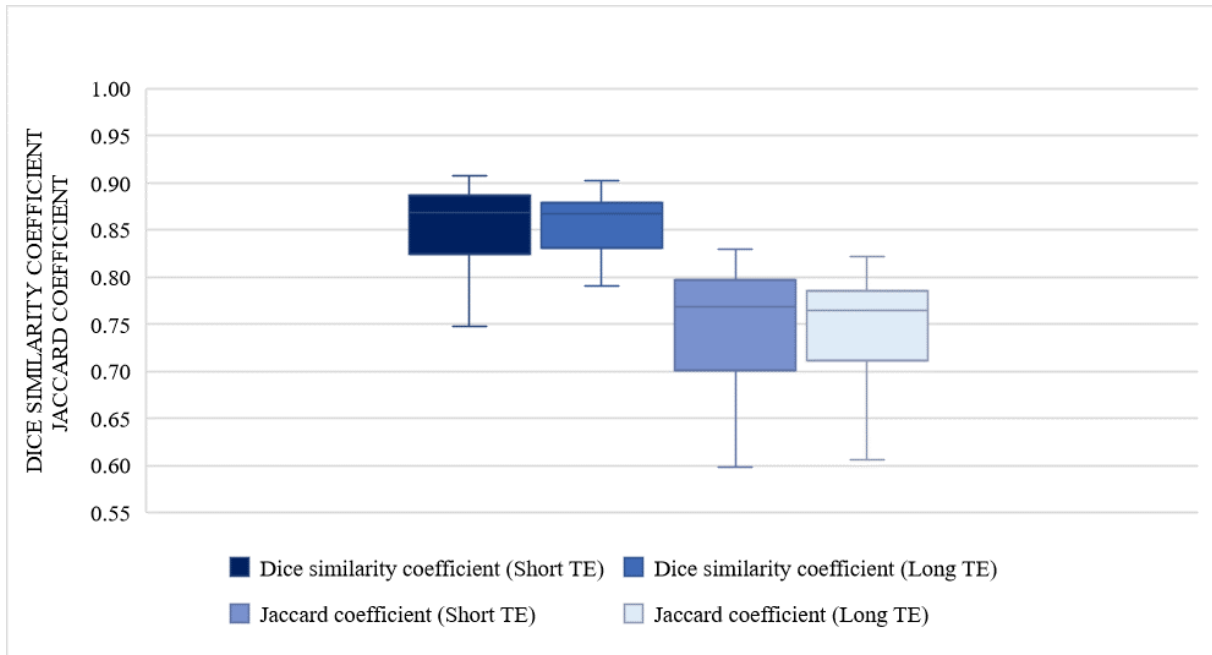


Figure 7. Box (containing the median, upper and lower quartile) and whisker (showing the range) plot based on DSC and Jaccard coefficient for all individuals for registration with T2_{raw,11.1}–images (short TE) and T2_{raw,88.8}–images (long TE).

3.2 Influence of mask on the registration quality

No visual difference could be distinguished in the quality of the registration between registrations with and without mask (Figure 8). Calculation time for registration was measured to 1 minute and 16 seconds, both for registration without mask and mask applied.



Figure 8. Typical image for unloaded spine (left), deformable registered image with no mask applied (middle) and deformable registered image with a rectangular mask applied (right).

The DSC and the Jaccard coefficient for all included individuals with the mask applied were 0.845 ± 0.057 [0.676-0.900] and 0.736 ± 0.081 [0.511-0.819], respectively (Figure 9). The mean of the Jacobian determinant for the individuals was 1.005 ± 0.032 [0.924-1.066].

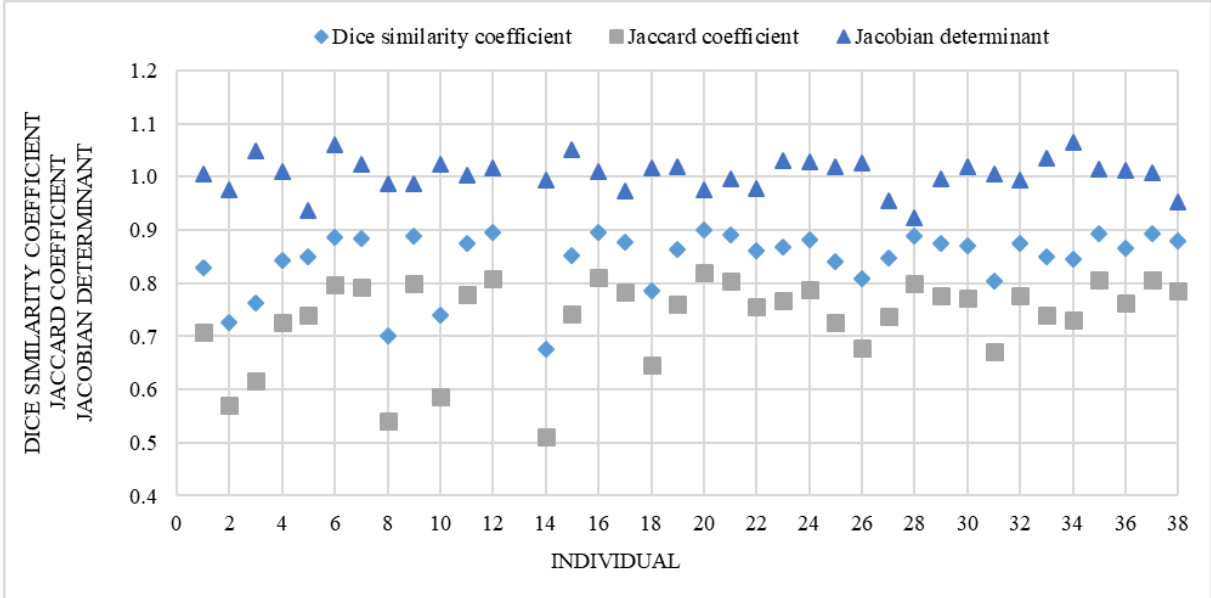


Figure 9. The DSC, Jaccard coefficient and mean of Jacobian determinant for the included individuals using three slices with mask applied.

The mean, median, first and third quartile of DSC and Jaccard coefficient (Figure 10) for masked image were 0.845, 0.865, 0.842 and 0.884 and 0.736, 0.763, 0.726 and 0.792, respectively. The corresponding values for the Jacobian determinant (Figure 10) for registration with whole image and registration with mask applied were, 1.035, 1.029, 1.001, 1.060 and 1.005, 1.009, 0.998, 1.024, respectively. No statistically significant difference between registration with or without mask was determined ($p = 0.247$). Also, the TOST test showed equivalence at 95% confidence interval.

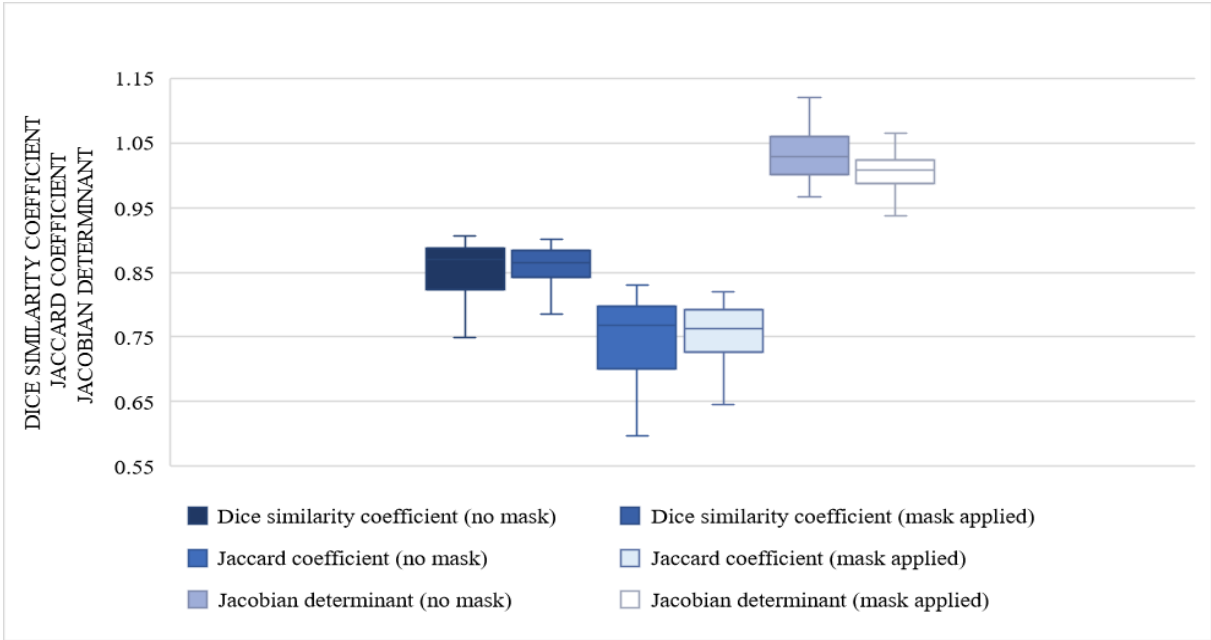


Figure 10. Box (containing the median, upper and lower quartile) and whisker (showing the range) plot based on DSC, Jaccard coefficient and Jacobian determinant for the included individuals for registration done with full image and registration with mask applied.

3.3 Optimization of the non-rigid registration step

As shown in Table 2, the DSC and Jaccard coefficient increased with a lower value of the final B-spline grid spacing, and as well as the computation time. A third-order of the B-spline interpolation showed lower DSC and Jaccard coefficient.

Table 2. The parameters FinalGridSpacingInVoxels and BSplineInterpolationOrder was changed according to the table for optimization of the non-rigid registration and presented in DSC, Jaccard coefficient and computation time.

| FinalGridSpacingInVoxels | Dice similarity coefficient | Jaccard coefficient | Time |
|---------------------------|-----------------------------|---------------------|------------|
| 4 | 0.824 | 0.701 | 8 min 55 s |
| 6 | 0.813 | 0.685 | 3 min 29 s |
| 8 | 0.806 | 0.676 | 3 min 15 s |
| 16 (default) | 0.798 | 0.664 | 2 min 45 s |
| 24 | 0.798 | 0.663 | 2 min 42 s |
| 32 | 0.795 | 0.660 | 2 min 44 s |
| BSplineInterpolationOrder | | | |
| 1 (default) | 0.798 | 0.664 | 2 min 45 s |
| 3 | 0.794 | 0.659 | 2 min 43 s |

3.4 Repeatability

For the IVD segmentation, the inter- and intra-observer agreement was excellent (Table 3). The ICC was 0.994 for DSC and 0.978 for the Jaccard coefficient.

Table 3. The DSC and Jaccard coefficient for six of the individuals that were reanalyzed for ICC determination.

| Individual | Dice similarity coefficient (1 st time segmentation) | Dice similarity coefficient (2 nd time segmentation) | Jaccard coefficient (1 st time segmentation) | Jaccard coefficient (2 nd time segmentation) |
|------------|---|---|---|---|
| 1 | 0.798 | 0.799 | 0.664 | 0.666 |
| 2 | 0.761 | 0.764 | 0.615 | 0.618 |
| 3 | 0.764 | 0.752 | 0.537 | 0.602 |
| 4 | 0.854 | 0.846 | 0.745 | 0.733 |
| 5 | 0.850 | 0.845 | 0.740 | 0.733 |
| 6 | 0.881 | 0.876 | 0.787 | 0.780 |

4 Discussion

To enable high performance registration and at the same time enable detailed evaluation of loading effects on pixel by pixel basis, an optimized imaging strategy has been developed, where short echo-time raw data images for the T2-mapping scanning were tuned to give a more T1-weighted contrast and the long echo-time raw data images to give more T2-weighted contrast.

There was no significant difference in DSC on a group level between registrations on short and long echo-time raw data images. That is, no clear trend towards higher or lower DSC between short and long TE was seen. Hence, the registration can be performed on both short and long echo-time raw data images and the proposed registration strategy for future detailed characterization of disc loading behavior is feasible. Further studies on how functional measures are affected by the registration strategy remains to be investigated.

Although DSC did not differ at group level, these estimates differed widely in some individuals. The detected difference in registration between short and long echo-time images in these individuals was probably not associated with the segmentation process. The same segmented ROI was used both in the evaluation of the short and long echo-time images. However, the contrast differed between these images and this may have affected the performance of the signal-intensity based registration. Hence, if Elastix is used for image registration, the images should have similar contrast. Preferably, the images should be acquired with a contrast that enhances the visibility of the tissue towards the background for improved segmentation and, for the purpose of LBP diagnosis, we encourage registration of images using short echo-time raw data images with a flatter IVD contrast to avoid registration of signal intensity changes within the IVD due to loading. With such strategy, the loading behavior of the IVD can be preserved throughout the registration process and the resulting registration can be transformed to the corresponding T2-maps for detailed characterization of the loading behavior. Future studies are planned to confirm the preservation of the IVD loading effect using the suggested registration strategy.

The present study investigated if the registration strategy benefited from masking and parameter optimization. The MR images were limited by a binary mask to improve the computational efficiency but also to improve the quality of the registration by excluding regions with non-relevant complex anatomy. For example, colon and other organs in the abdomen was excluded which can move or contain air and this could affect the quality of the registration, resulting in a lower DSC as well as negative Jacobian determinants. Elastix registration is based on local signal intensity and thus variations in signal intensity in background far from disc should not affect. As shown by our findings, no visual difference could be distinguished in the quality between registration with and without mask. Also, masking did not significantly improve the registration quality as measured by the DSC. Hence, the initial registration set-up, using the whole image, was considered good. Moreover, masking did not speed up the registration process and, since it takes time to create a mask, we do not recommend using a mask for the registration strategy. The mean of the Jacobian determinant was slightly different for registrations with mask than for registrations without mask. Since the mean of the Jacobian determinant was determined for the entire image and, hence, reflects also the registration of the background with air filled colon in motion rather than the area of interest, a higher mean Jacobian determinant should be obtained due to the general expansion of points in the background during the deformation (Figure 10).

Two parameters were varied to see if the quality of the registration could be improved. The parameter `FinalGridSpacingInVoxels` determine the size of the B-spline grid in the last resolution level, which could affect the accuracy based on the size selected. However, the initially high registration quality was only slightly improved by lowering the final B-Spline grid spacing at the expense of increased computation time. Since the initial accuracy, using a default setting of `FinalGridSpacingInVoxels = 16`, was sufficiently high with a reasonable computation time, we came to the conclusion that there is no need to decrease the value of this parameter when registration is performed using the proposed strategy. The second parameter changed was the `BSplineInterpolationOrder`, which was set to third-order but showed no remarkable difference in neither accuracy nor time from the first-order. As mentioned in Elastix manual, a setting of third-order may improve the accuracy. However, since the accuracy of the registration already was high, changing it from first-order did not affect the result. For more complex structures, accuracy can possibly be improved with third-order interpolation.

In agreement with the previous study [17], the agreement between repeated segmentations was excellent using the present segmentation software. This implies that the sensitivity for observer's dependency in the quality measures was low. The actual observer had little experience of spinal images, still the first- and the second segmentation were similar and high repeatability was proven. Also, comparing registration using segmentation with nine slices versus three slices gave a negligible mean difference in the DSC value, demonstrating the performance of the registration, as well as the segmentation. The fact that the DSC were low for some IVDs may be associated with the size of the IVD. Since the IVD is a small structure with only a few voxels, the measures are very sensitive to the inclusion or exclusion of single voxels in the segmentation. This in combination with the difficulty in depicting the contrast edge of the IVD makes the segmentation challenging. Automatic segmentation methods with for example deep learning is probably a viable solution to reduce this problem.

For five of the individuals, the DSC values were slightly lower. Two of the individuals (ID2 and ID3) had moved during imaging in the loaded state and caused movement artifacts (see Appendix for images). This may have compromised the segmentation. In another individual (ID8), the complex shape of the IVDs made the segmentation of the contour extra challenging. The size of the disc under L5 in unloaded state differ greatly from the disc in loaded state for another individual (ID10). For the last individual (ID14), the position of the IVD differed greatly in the unloaded and loaded spine image, *e.g.* the disc above L5 was placed higher up in unloaded imaging relative to loaded image (see Appendix for images). All these effects may have had a compromising effect on the registration.

Limitations

In this study, the quality of the registration was assessed using DSC and Jaccard coefficient. These metrics are limited in sensitivity for high quality registrations as the measures are based on reliable segmentation and as such are dependent on the performance of the observer. Also, the IVD is a small structure with only a few voxels. This makes the measure very sensitive to the inclusion or exclusion of a single voxel in the segmentation. Future work is encouraged to develop better measures for evaluation of registration quality, preferably not limited to the segmentation procedure.

5 Conclusion

In this study, a new promising strategy for registration of spine MR images with Elastix was developed. The registration showed high quality evaluated with DSC and Jaccard coefficient. The evaluation of the Jacobian determinant indicated a preserved topology of the deformation. There was no significant difference between registration with $T2_{\text{raw},11.1}$ -images and $T2_{\text{raw},88.8}$ -images, which indicates that the strategy is promising for extraction of detailed functional image markers. Also, neither did the mask have an impact on the quality of the registration. Deformable image registration of spine MR images with the optimized registration strategy seems feasible and useful for detailed characterization of disc loading behaviours.

Reference list

1. Hansen Brandt, B., Hansen, P., Carrino, John A., Fournier, G., Rasti, Z., & Boesen, M. (2016). Imaging in mechanical back pain: Anything new? *Best Practice & Research Clinical Rheumatology*, 30(4), 766-785.
2. Hebelka, H., Torén, L., Lagerstrand, K., & Brisby, H. (2018). Axial loading during MRI reveals deviant characteristics within posterior IVD regions between low back pain patients and controls. *European Spine Journal*, 2018, Vol. 27, Iss. 11, Pp. 2840-2846, 27(11), 2840-2846.
3. Peng, B., Wu, W., Hou, S., Li, P., Zhang, C., & Yang, Y. (2005). The pathogenesis of discogenic low back pain. *The Journal of Bone and Joint Surgery. British Volume*, 87(1), 62-7.
4. Maher, C., Underwood, M., & Buchbinder, R. (2017). Non-specific low back pain. *The Lancet*, 389(10070), 736-747.
5. Hebelka, H., & University of Gothenburg. (2014). Discogenic pain – A diagnostic challenge. [Department of Orthopaedics, Institute of Clinical Sciences at Sahlgrenska Academy, University of Gothenburg].
6. Andersson, J., & KTH (2018). T2 Mapping compared to standard MRI assessment – An assessment of the knee cartilage on distal femur. [KTH Royal Institute of Technology School of Engineering Sciences in Chemistry, Biotechnology and Health].
7. Antoniou, J., Pike, G., Steffen, T., Baramki, H., Poole, A., Aebi, M., & Alini, M. (1998). Quantitative magnetic resonance imaging in the assessment of degenerative disc disease. *Magnetic Resonance in Medicine*, 40(6), 900-907.
8. Mwale, F., Iatridis, J., & Antoniou, C. (2008). Quantitative MRI as a diagnostic tool of intervertebral disc matrix composition and integrity. *European Spine Journal*, 17(Supplement 4), 432-440.
9. Nilsson, M., Lagerstrand, K., Kasperska, I., Brisby, H., & Hebelka, H. (2016). Axial loading during MRI influences T2-mapping values of lumbar discs: A feasibility study on patients with low back pain. *European Spine Journal*, 2016, Vol. 25, Iss. 9, Pp. 2856-2863, 25(9), 2856-2863.
10. Nazib, A., Galloway, J., Fookes, C., & Perrin, D. (2018). Performance of Image Registration Tools on High-Resolution 3D Brain Images. *ArXiv.org*, ArXiv.org, Jul 13, 2018.
11. Malinsky, Milos, Peter, Roman, Hodneland, Erlend, Lundervold, Astri, Lundervold, Arvid, & Jan, Jiri. (2013). Registration of FA and T1-Weighted MRI Data of Healthy Human Brain Based on Template Matching and Normalized Cross-Correlation. *Journal of Digital Imaging*, 26(4), 774-785.

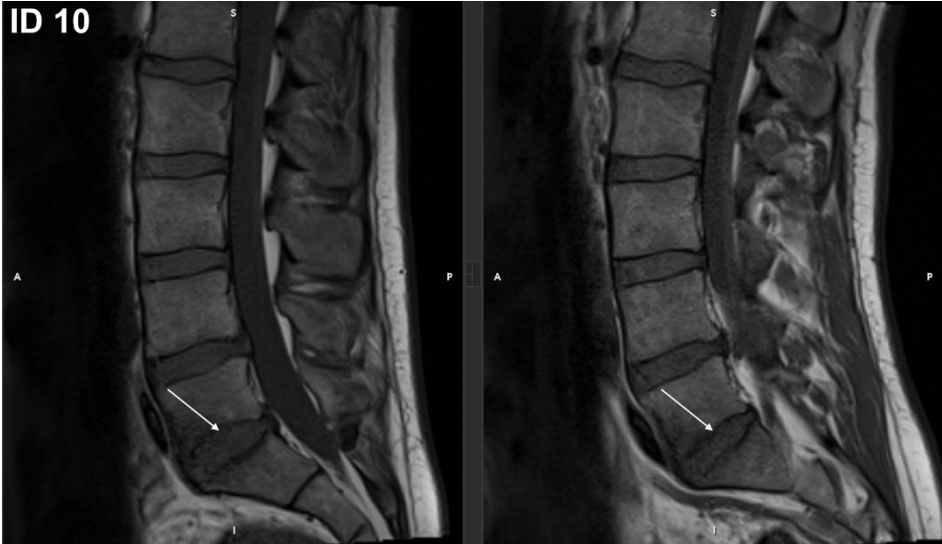
12. Klein, S., Staring, M., Murphy, K., Viergever, M., & Pluim, J. (2010). Elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*, 29(1), 196-205.
13. Oliveira, F., & Tavares, J. (2014). Medical image registration: A review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2), 73-93.
14. Song, J., Christensen, Gary E., Durumeric, Oguz, Johnson, Hans, Kuhl, Jon, & Saha, Punam. (2017). *Methods for Evaluating Image Registration*, ProQuest Dissertations and Theses.
15. Elastix, <https://github.com/SuperElastix/elastix/wiki/What-is-elastix> (April 12th 2020)
16. Elastix, <https://github.com/SuperElastix/elastix/wiki> (May 9th 2020)
17. Waldenberg, C., Hebelka, H., Brisby, H., & Lagerstrand, K. (2019). Differences in IVD characteristics between low back pain patients and controls associated with HIZ as revealed with quantitative MRI. *Plos One*, 2019, Vol. 14, Iss. 8, 14(8), Plos One, 2019, Vol. 14, Iss. 8.
18. Klein S., & Staring M. (2019). Elastix, The manual
19. Koo, T., & Li, M. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155-163.

Appendix A

Spine images of five individuals with low DSC and Jaccard coefficient



ID 10



ID14



Appendix B

Parameter file for default settings in Elastix for rigid registration

// Example parameter file for rotation registration

// C-style comments: //

// The internal pixel type, used for internal computations

// Leave to float in general.

// NB: this is not the type of the input images! The pixel

// type of the input images is automatically read from the

// images themselves.

// This setting can be changed to "short" to save some memory

// in case of very large 3D images.

(FixedInternalImagePixelType "float")

(MovingInternalImagePixelType "float")

// The dimensions of the fixed and moving image

// Up to elastix 4.5 this had to be specified by the user.

// From elastix 4.6, this is not necessary anymore.

// (FixedImageDimension 2)

// (MovingImageDimension 2)

// Specify whether you want to take into account the so-called

// direction cosines of the images. Recommended: true.

// In some cases, the direction cosines of the image are corrupt,

// due to image format conversions for example. In that case, you

// may want to set this option to "false".

(UseDirectionCosines "true")

// ***** Main Components *****

// The following components should usually be left as they are:

(Registration "MultiResolutionRegistration")

(Interpolator "BSplineInterpolator")

(ResampleInterpolator "FinalBSplineInterpolator")

(Resampler "DefaultResampler")

// These may be changed to Fixed/MovingSmoothingImagePyramid.

// See the manual.

(FixedImagePyramid "FixedRecursiveImagePyramid")

(MovingImagePyramid "MovingRecursiveImagePyramid")

// The following components are most important:

// The optimizer AdaptiveStochasticGradientDescent (ASGD) works

// quite ok in general. The Transform and Metric are important

// and need to be chosen careful for each application. See manual.

(Optimizer "AdaptiveStochasticGradientDescent")

(Transform "EulerTransform")

(Metric "AdvancedMattesMutualInformation")

// ***** Transformation *****

// Scales the rotations compared to the translations, to make

// sure they are in the same range. In general, it's best to

// use automatic scales estimation:

(AutomaticScalesEstimation "true")

// Automatically guess an initial translation by aligning the

// geometric centers of the fixed and moving.

(AutomaticTransformInitialization "true")

// Whether transforms are combined by composition or by addition.

// In generally, Compose is the best option in most cases.

// It does not influence the results very much.

(HowToCombineTransforms "Compose")

// ***** Similarity measure *****

// Number of grey level bins in each resolution level,

// for the mutual information. 16 or 32 usually works fine.

```

// You could also employ a hierarchical strategy:
// (NumberOfHistogramBins 16 32 64)
(NumberOfHistogramBins 32)

// If you use a mask, this option is important.
// If the mask serves as region of interest, set it to false.
// If the mask indicates which pixels are valid, then set it to true.
// If you do not use a mask, the option doesn't matter.
(ErodeMask "false")

// ***** Multiresolution *****
// The number of resolutions. 1 Is only enough if the expected
// deformations are small. 3 or 4 mostly works fine. For large
// images and large deformations, 5 or 6 may even be useful.
(NumberOfResolutions 4)

// The downsampling/blurring factors for the image pyramids.
// By default, the images are downsampled by a factor of 2
// compared to the next resolution.
// So, in 2D, with 4 resolutions, the following schedule is used:
// (ImagePyramidSchedule 8 8 4 4 2 2 1 1 )
// And in 3D:
// (ImagePyramidSchedule 8 8 8 4 4 4 2 2 2 1 1 1 )
// You can specify any schedule, for example:
// (ImagePyramidSchedule 4 4 4 3 2 1 1 1 )
// Make sure that the number of elements equals the number
// of resolutions times the image dimension.

// ***** Optimizer *****
// Maximum number of iterations in each resolution level:
// 200-500 works usually fine for rigid registration.
// For more robustness, you may increase this to 1000-2000.
(MaximumNumberOfIterations 250)

```

```

// The step size of the optimizer, in mm. By default the voxel size is used.
// which usually works well. In case of unusual high-resolution images
// (eg histology) it is necessary to increase this value a bit, to the size
// of the "smallest visible structure" in the image:
// (MaximumStepLength 1.0)

// ***** Image sampling *****
// Number of spatial samples used to compute the mutual
// information (and its derivative) in each iteration.
// With an AdaptiveStochasticGradientDescent optimizer,
// in combination with the two options below, around 2000
// samples may already suffice.
// (NumberOfSpatialSamples 2048)

// Refresh these spatial samples in every iteration, and select
// them randomly. See the manual for information on other sampling
// strategies.
// (NewSamplesEveryIteration "true")
// (ImageSampler "Random")

// ***** Interpolation and Resampling *****
// Order of B-Spline interpolation used during registration/optimisation.
// It may improve accuracy if you set this to 3. Never use 0.
// An order of 1 gives linear interpolation. This is in most
// applications a good choice.
// (BSplineInterpolationOrder 1)

// Order of B-Spline interpolation used for applying the final
// deformation.
// 3 gives good accuracy; recommended in most cases.
// 1 gives worse accuracy (linear interpolation)
// 0 gives worst accuracy, but is appropriate for binary images

```



```
// (masks, segmentations); equivalent to nearest neighbor interpolation.  
(FinalBSplineInterpolationOrder 3)
```

```
//Default pixel value for pixels that come from outside the picture:  
(DefaultPixelValue 0)
```

```
// Choose whether to generate the deformed moving image.  
// You can save some time by setting this to false, if you are  
// only interested in the final (nonrigidly) deformed moving image  
// for example.  
(WriteResultImage "true")
```

```
// The pixel type and format of the resulting deformed moving image  
(ResultImagePixelFormat "short")  
(ResultImageFormat "mhd")
```

Parameter file for default settings in Elastix for non-rigid registration

// Example parameter file for B-spline registration

// C-style comments: //

// The internal pixel type, used for internal computations

// Leave to float in general.

// NB: this is not the type of the input images! The pixel

// type of the input images is automatically read from the

// images themselves.

// This setting can be changed to "short" to save some memory

// in case of very large 3D images.

(FixedInternalImagePixelType "float")

(MovingInternalImagePixelType "float")

// The dimensions of the fixed and moving image

// Up to elastix 4.5 this had to be specified by the user.

// From elastix 4.6, this is not necessary anymore.

// (FixedImageDimension 2)

// (MovingImageDimension 2)

// Specify whether you want to take into account the so-called

// direction cosines of the images. Recommended: true.

// In some cases, the direction cosines of the image are corrupt,

// due to image format conversions for example. In that case, you

// may want to set this option to "false".

(UseDirectionCosines "true")

// ***** Main Components *****

// The following components should usually be left as they are:

(Registration "MultiResolutionRegistration")

(Interpolator "BSplineInterpolator")

(ResampleInterpolator "FinalBSplineInterpolator")

(Resampler "DefaultResampler")

```

// These may be changed to Fixed/MovingSmoothingImagePyramid.
// See the manual.
(FixedImagePyramid "FixedRecursiveImagePyramid")
(MovingImagePyramid "MovingRecursiveImagePyramid")

// The following components are most important:
// The optimizer AdaptiveStochasticGradientDescent (ASGD) works
// quite ok in general. The Transform and Metric are important
// and need to be chosen careful for each application. See manual.
(Optimizer "AdaptiveStochasticGradientDescent")
(Transform "BSplineTransform")
(Metric "AdvancedMattesMutualInformation")

// ***** Transformation *****
// The control point spacing of the bspline transformation in
// the finest resolution level. Can be specified for each
// dimension differently. Unit: mm.
// The lower this value, the more flexible the deformation.
// Low values may improve the accuracy, but may also cause
// unrealistic deformations. This is a very important setting!
// We recommend tuning it for every specific application. It is
// difficult to come up with a good 'default' value.
// (FinalGridSpacingInPhysicalUnits 16)

// Alternatively, the grid spacing can be specified in voxel units.
// To do that, uncomment the following line and comment/remove
// the FinalGridSpacingInPhysicalUnits definition.
(FinalGridSpacingInVoxels 16)

// By default the grid spacing is halved after every resolution,
// such that the final grid spacing is obtained in the last
// resolution level. You can also specify your own schedule,

```

```

// if you uncomment the following line:
// (GridSpacingSchedule 4.0 4.0 2.0 1.0)
// This setting can also be supplied per dimension.

// Whether transforms are combined by composition or by addition.
// In generally, compose is the best option in most cases.
// It does not influence the results very much.
(HowToCombineTransforms "Compose")

// ***** Similarity measure *****
// Number of grey level bins in each resolution level,
// for the mutual information. 16 or 32 usually works fine.
// You could also employ a hierarchical strategy:
// (NumberOfHistogramBins 16 32 64)
(NumberOfHistogramBins 32)

// If you use a mask, this option is important.
// If the mask serves as region of interest, set it to false.
// If the mask indicates which pixels are valid, then set it to true.
// If you do not use a mask, the option doesn't matter.
(ErodeMask "false")

// ***** Multiresolution *****
// The number of resolutions. 1 Is only enough if the expected
// deformations are small. 3 or 4 mostly works fine. For large
// images and large deformations, 5 or 6 may even be useful.
(NumberOfResolutions 4)

// The downsampling/blurring factors for the image pyramids.
// By default, the images are downsampled by a factor of 2
// compared to the next resolution.
// So, in 2D, with 4 resolutions, the following schedule is used:
// (ImagePyramidSchedule 8 8 4 4 2 2 1 1 )

```

```

// And in 3D:
// (ImagePyramidSchedule 8 8 8 4 4 4 2 2 2 1 1 1 )
// You can specify any schedule, for example:
// (ImagePyramidSchedule 4 4 4 3 2 1 1 1 )
// Make sure that the number of elements equals the number
// of resolutions times the image dimension.

// ***** Optimizer *****
// Maximum number of iterations in each resolution level:
// 200-2000 works usually fine for nonrigid registration.
// The more, the better, but the longer computation time.
// This is an important parameter!
(MaximumNumberOfIterations 500)

// The step size of the optimizer, in mm. By default the voxel size is used.
// which usually works well. In case of unusual high-resolution images
// (eg histology) it is necessary to increase this value a bit, to the size
// of the "smallest visible structure" in the image:
// (MaximumStepLength 1.0)

// ***** Image sampling *****
// Number of spatial samples used to compute the mutual
// information (and its derivative) in each iteration.
// With an AdaptiveStochasticGradientDescent optimizer,
// in combination with the two options below, around 2000
// samples may already suffice.
(NumberOfSpatialSamples 2048)

// Refresh these spatial samples in every iteration, and select
// them randomly. See the manual for information on other sampling
// strategies.
(NewSamplesEveryIteration "true")
(ImageSampler "Random")

```

```
// ***** Interpolation and Resampling *****
// Order of B-Spline interpolation used during registration/optimisation.
// It may improve accuracy if you set this to 3. Never use 0.
// An order of 1 gives linear interpolation. This is in most
// applications a good choice.
(BSplineInterpolationOrder 1)

// Order of B-Spline interpolation used for applying the final
// deformation.
// 3 gives good accuracy; recommended in most cases.
// 1 gives worse accuracy (linear interpolation)
// 0 gives worst accuracy, but is appropriate for binary images
// (masks, segmentations); equivalent to nearest neighbor interpolation.
(FinalBSplineInterpolationOrder 3)

//Default pixel value for pixels that come from outside the picture:
(DefaultPixelValue 0)

// Choose whether to generate the deformed moving image.
// You can save some time by setting this to false, if you are
// not interested in the final deformed moving image, but only
// want to analyze the deformation field for example.
(WriteResultImage "true")

// The pixel type and format of the resulting deformed moving image
(ResultImagePixelFormat "short")
(ResultImageFormat "mhd")
```