IOP Publishing Phys. Med. Biol. 65 (2020) 015006 (19pp) https://doi.org/10.1088/1361-6560/ab501d

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

RECEIVED
12 July 2019

REVISED
2 October 2019

ACCEPTED FOR PUBLICATION
22 October 2019

PUBLISHED
10 January 2020

Biomechanical quality assurance criteria for deformable image registration algorithms used in radiotherapy guidance

Cornel Zachiu[1,4], Baudouin Denis de Senneville[2,3], Bas W Raaymakers[1] and Mario Ries[2]

[1] Department of Radiotherapy, UMC Utrecht, Heidelberglaan 100, 3584 CX, The Netherlands
[2] Imaging Division, UMC Utrecht, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands
[3] Institut de Mathématiques de Bordeaux (IMB), UMR 5251 CNRS/University of Bordeaux, F-33400 Talence, France
[4] Author to whom correspondence should be addressed.

E-mail: C.Zachiu@umcutrecht.nl

Keywords: deformable image registration, quality assurance, biomechanics

Abstract

Image-guided radiation therapy (IGRT) allows radiation dose deposition with a high degree of geometric accuracy. Previous studies have demonstrated that such therapies may benefit from the employment of deformable image registration (DIR) algorithms, which allow both the automatic tracking of anatomical changes and accumulation of the delivered radiation dose over time. In order to ensure patient care and safety, however, the estimated deformations must be subjected to stringent quality assurance (QA) measures. In the present study we propose to extend the state-of-the-art methodology for QA of DIR algorithms by a set of novel biomechanical criteria.

The proposed biomechanical criteria imply the calculation of the normal and shear mechanical stress, which would occur within the observed tissues as a result of the estimated deformations. The calculated stress is then compared to plausible physiological limits, providing thus the anatomical plausibility of the estimated deformations. The criteria were employed for the QA of three DIR algorithms in the context of abdominal conebeam computed tomography and magnetic resonance radiotherapy guidance.

An initial evaluation of organ boundary alignment capabilities indicated that all three algorithms perform similarly. However, an analysis of the deformations within the organ boundaries with respect to the proposed biomechanical QA criteria revealed different degrees of anatomical plausibility. Additionally, it was demonstrated that violations of these criteria are also indicative of errors within the dose accumulation process.

The proposed QA criteria, therefore, provide a tissue-dependent assessment of the anatomical plausibility of the deformations estimated by DIR algorithms, showcasing potential in ensuring patient safety for future adaptive IGRT treatments.

1. Introduction

Image guided external beam radiotherapy (IGRT) allows the deposition of the therapeutic radiation dose with a high degree of geometric accuracy at the site of the tumor, while at the same time maximizing healthy tissue sparing (Roach *et al* 2011, Jaffray 2012). With the recent integration of imaging systems such as cone beam computed tomography (CBCT) (Guckenberger 2011) or magnetic resonance imaging (MRI) (Raaymakers *et al* 2017) into radiotherapy machines, image guidance of radiotherapy treatments plays a role in all major phases of the treatment including diagnosis, planning, response assessment and delivery of the therapeutic segments. In particular, image guidance allows visualizing anatomical changes induced during treatment by physiological motion, which has been previously identified to be a major source of uncertainties during external beam radiotherapy (Bortfeld *et al* 2004, Keall *et al* 2006, Korreman 2012). In this sense, deformable image registration (DIR) algorithms have demonstrated great potential in automatically tracking such changes, having at the same time the capability of delivering a dose deposition history in a spatially consistent manner (i.e. dose accumulation) (Brock *et al* 2017, Oh and Kim 2017). This in turn allows potential adjustments to the treatment

plan with respect to the physiological motion (Kontaxis *et al* 2015, 2017). It is therefore imperative for both the optimal delivery of IGRT and, more importantly, patient safety to subject the deformations estimated by the employed DIR algorithm(s) to stringent quality assurance (QA) measures.

A set of recommended performance indicators and QA solutions for DIR algorithms used in IGRT, are detailed within the AAPM report of Task Group 132 (Brock *et al* 2017). Some of the most representative include the target registration error (TRE) (Maurer *et al* 1997), the dice similarity coefficient (DSC) (Dice 1945) and the mean distance to agreement (MDA) (Chalana and Kim 1997). Such criteria are, however, suitable for evaluating DIR performance within high-contrast image areas such as organ boundaries and/or fiducial markers, lacking the necessary sensitivity within homogeneous image regions. A previously proposed solution in this sense is the usage of physical or software phantoms, where a set of known deformations is applied to an object or an image and subsequently used as a gold standard to assess the accuracy and precision of registration algorithms (Brock *et al* 2017). However, both the usage of phantoms and the aforementioned high-contrast criteria are generally employed as prospective evaluation methods for DIR algorithms, with the assumption that they will maintain their accuracy and precision when employed in a clinical or interventional setting (Liu *et al* 2012, Pluim *et al* 2016). This can become problematic, since in such instances, unforeseen image artifacts, distortions and noise may affect algorithm performance.

Current clinical practice for evaluation of DIR results in an interventional setting implies visual or (semi-) automatic assessment of organ boundary or marker (anatomical or implanted) alignment (Bissonnette *et al* 2012, Brock *et al* 2017, Oh and Kim 2017). Aside visual inspection being subjective and error-prone, this does not guarantee that within contrast-devoid image regions the estimated deformations are accurate. While for automatic contour propagation misregistrations within the organ boundaries may not be an important issue, in case of mapping quantitative information such as radiation dose or Houndsfield units, it may have a serious impact on the adaptation of the therapeutic plan and implicitly patient safety and treatment. Existing solutions suggest the analysis of deformation-dependent mathematical functions such as the Jacobian determinant and the curl in order to determine the anatomical plausibility of deformations estimated within soft tissue boundaries (Schreibmann *et al* 2012, Sotiras *et al* 2013, Brock *et al* 2017, Zachiu *et al* 2018). Due to their high water/blood volume, such tissues are near-incompressible and in effect the Jacobian determinant of their deformations should be close to one (Rohlfing *et al* 2003, Bistoquet *et al* 2008, Zachiu *et al* 2018). Thus, high deviations of the Jacobian determinant from one in soft-tissue structures such as liver, kidney [...] are anatomically implausible. Similarly, during typical physiological motion, strong soft tissue torsions are not expected within the organ boundaries and therefore the magnitude of the curl of the estimated deformations is expected to be close to zero (Schreibmann *et al* 2012, Zachiu *et al* 2018). An assessment of anatomical plausibility based on the Jacobian determinant and the curl magnitude is however, rather generic and of qualitative value. This is due to the fact that, depending on their material properties, biological soft tissues do indeed allow a particular and individual degree of compression/expansion and torsion.

The contribution of the current study is twofold:

(i) We propose an extension to the existing methodology for QA of DIR of biological soft tissues by a set of tissue-specific QA criteria which take into account the particular biomechanical properties of the observed anatomical structures. As a principle, the criteria evaluate the mechanical stress which would result within the tissues as an effect of the estimated deformations, which is subsequently compared against non-physiological thresholds.

(ii) The applicability of the proposed biomechanical QA criteria is demonstrated for two clinically-relevant IGRT scenarios: that of CBCT and that of MR guidance. Additionally, the correspondence between violations of the proposed criteria and errors within a radiation dose distribution mapped via estimated deformations is also investigated.

2. Method description

2.1. Description of the employed QA criteria

2.1.1. Existing QA criteria

Within the scope of this work, several existing QA criteria for deformable image registration were selected to complement the proposed criteria relying on biomechanical tissue properties (which will be described shortly in section 2.1.2).

(1) One of the current clinical standards for validating registration results in an interventional setting consists in a visual inspection of post-registration alignment of organ boundaries. As a representative objective criterion for such an approach, the current study employs the Dice Similarity Coefficient (DSC). The DSC can be used to quantify the overlap between two contours and is mathematically defined as (Dice 1945):

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{1}$$

where A and B are the two contours, $A \cap B$ is their intersection and $|\cdot|$ denotes the cardinality of a set (i.e. the number of voxels).

(2) As specified within the introduction, the applicability of the DSC as a QA measure within contrast-devoid regions is limited. In effect, as previous studies suggest, here we also analyze the Jacobian determinant and the curl magnitude of the deformations as QA criteria in such areas. The two functions are defined as:

$$J(\vec{r} + \mathbf{u}(\vec{r})) = \begin{vmatrix} 1 + \frac{\partial u_1(\vec{r})}{\partial x} & \frac{\partial u_1(\vec{r})}{\partial y} & \frac{\partial u_1(\vec{r})}{\partial z} \\ \frac{\partial u_2(\vec{r})}{\partial x} & 1 + \frac{\partial u_2(\vec{r})}{\partial y} & \frac{\partial u_2(\vec{r})}{\partial z} \\ \frac{\partial u_3(\vec{r})}{\partial x} & \frac{\partial u_3(\vec{r})}{\partial y} & 1 + \frac{\partial u_3(\vec{r})}{\partial z} \end{vmatrix} \tag{2}$$

$$\mathrm{curl}(\vec{\mathbf{u}}(\vec{r})) = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ u_1(\vec{r}) & u_2(\vec{r}) & u_3(\vec{r}) \end{vmatrix}. \tag{3}$$

In particular, as argued in the introduction, deviations from one and zero of the Jacobian determinant and curl magnitude, respectively, will be investigated.

2.1.2. Proposed biomechanical QA criteria for deformable image registration

In this study, we propose three voxel-wise biomechanical QA criteria, which assist in evaluating the anatomical plausibility of the deformations estimated by DIR algorithms. Note that the criteria are constructed under the assumption that the tissues under observation follow a linear elastic model (Doyley and Parker 2014).

(1) The first proposed biomechanical criterion is related to the tensile/compressive stress to which a voxel of tissue would be subjected as a result of the deformations estimated within the voxel. It is proportional to the volume change induced by the deformations, and is calculated via the following equation (Hostettler *et al* 2010):

$$\sigma(\vec{r}) = -\frac{E(\vec{r})}{3(1 - 2\nu(\vec{r}))}\left(\frac{\Delta V(\vec{r})}{V(\vec{r})}\right) \tag{4}$$

where E and ν are the elastic modulus and Poisson ratio of the tissue, ΔV and V are the volume change and the initial volume of the voxel, and $\vec{r}$ is the spatial location of the voxel within the image. Equation (4) can be re-written as:

$$\sigma(\vec{r}) = \frac{E(\vec{r})}{3(1 - 2\nu(\vec{r}))}(J(\vec{r}) - 1) \tag{5}$$

where $J(\vec{r})$ is the Jacobian determinant of the deformation estimated within a particular voxel (Rohlfing *et al* 2003). In the scope of this study, the tensile/compressive stress will be computed using equation (5).

When subjected to increasing tensile/compressive mechanical strain, biological tissues will typically experience increasing/decreasing mechanical stress up to a particular threshold beyond which the tissue undergoes plastic deformation. As a QA criterion for registration algorithms we propose verifying in each voxel whether the estimated deformation subjects the tissue to a tensile/compressive strain that would theoretically lead to exceeding this threshold.

(2) The stress calculated in equation (5) can also be interpreted as a change in pressure as a result of the voxel changing volume. In particular, for pressure induced by compressive strain, we propose as a second biomechanical QA criterion verifying whether this volume change results in a pressure gradient that exceeds that of arterial blood pressure. Such a criterion is justified by the fact that, during typical anatomical/physiological motion occurring in patients in a resting position during therapeutic imaging or in treatment position for external beam radiotherapy, internal organs are usually not subjected to a mechanical stress which would occlude normal blood circulation.

(3) The third biomechanical QA criterion proposed in the scope of this study relies on the analysis of the shear stress that would occur within the tissues as a result of the estimated deformations. Similar to the tensile/compressive stress, there is a limit of the shear stress beyond which the tissue undergoes deformation up to the points of tissue fracture/rupture. Thus, the third proposed biomechanical QA criterion checks whether the mechanical strain induced by the deformation estimated within a particular

voxel, leads to a plastic deformation as a result of shear stress. In the scope of this work, the shear stress is computed as (Landau *et al* 2012):

$$\gamma_{xy}(\vec{r}) = \frac{G(\vec{r})}{2}\left[\frac{\partial u_1(\vec{r})}{\partial y} + \frac{\partial u_2(\vec{r})}{\partial x}\right] \tag{6}$$

$$\gamma_{xz}(\vec{r}) = \frac{G(\vec{r})}{2}\left[\frac{\partial u_2(\vec{r})}{\partial z} + \frac{\partial u_3(\vec{r})}{\partial y}\right] \tag{7}$$

$$\gamma_{yz}(\vec{r}) = \frac{G(\vec{r})}{2}\left[\frac{\partial u_1(\vec{r})}{\partial z} + \frac{\partial u_3(\vec{r})}{\partial x}\right] \tag{8}$$

where $(\gamma_{xy}, \gamma_{xz}, \gamma_{yz})$ are the shear stresses in each plane of the 3D space, G is the shear modulus of the observed tissues and (u_1, u_2, u_3) are the 3D displacements estimated by the deformable registration algorithms. Note that for (near-) incompressible tissues the shear modulus can be approximated by $G \approx E/3$ (Landau *et al* 2012).

2.2. Experimental setup

In the scope of this study, two IGRT scenarios were considered: CBCT and MR guidance of radiotherapy, in hepatocellular carcinoma (HCC) patients. The experimental setup used to to demonstrate the applicability of the proposed QA criteria in such scenarios is schematically illustrated in figure 1. The setup was divided into four major components: input data, CSPAMM experiment, gold standard generation and QA of three selected DIR algorithms, all of which will be detailed in the following sections.

2.2.1. Input data

For the implementation of the proposed biomechanical QA criteria, the elastic modulus and the Poisson ratio of several tissues of interest were necessary (see equations (5)–(8)). In figure 1 these are referred to as 'elastic properties of tissues'. By the means of mechanical tests or US/MR elastography, previous studies have determined the elastic modulus for several tissues inside the human body. In the scope of the present work, the tissues of interest consist of: liver, HCC, kidney and ribs (cancellous bone), whose approximate elastic modulus is reported in table 1. The Poisson ratio (ν) for the soft tissues was considered to be 0.49 (Poisson ratio of near-incompressible materials), while for cancellous bone $\nu = 0.25$ (Lai *et al* 2015). Similarly, the tensile/compressive and shear stress leading to plastic deformations (indicated in figure 1 as 'biomechanical tissue thresholds (literature)') are tissue dependent and have been determined by previous studies mostly via mechanical tests. Table 2 reports the threshold values which were employed in the scope of this study. In addition, for the implementation of the third biomechanical criterion, the arterial blood pressure was considered to be 18 500 Pa (140 mmHg).

In terms of imaging data, a CT and a 3D MR image were selected from datasets acquired on hepatocellular carcinoma (HCC) patients. The CT image was used to generate a synthetic CBCT (sCBCT) using the TIGRE CBCT reconstruction toolbox (Feldkamp *et al* 1984, Biguri *et al* 2016). For both the sCBCT and the MR image, several anatomical structures of interest were delineated by an experienced staff member, including: liver, HCC, kidneys, ribs (only for the sCBCT image), lungs, spleen, bowels, body and connective adipose tissue. For a comparison of the effect of misregistration on dose accumulation for RT (described in a subsequent section of this paper), in both cases the intensity-modulated delivery of a single fraction of 20 Gy to the GTV was planned and simulated. The planning and simulation was performed using an in-house developed software described in Kontaxis *et al* (2017). Note that the delivery of a single fraction of 20 Gy was selected for demonstration purposes only. In a clinical setting, the prescribed dose for HCC is typically 70 Gy, which is delivered in a fractionated manner over the course of multiple therapy sessions. The sCBCT and MR images, their respective associated anatomical contours and the planned radiation distributions are referred to in figure 1 as 'sCBCT/MR anatomical images', 'sCBCT/MR organ contours' and 'sCBCT/MR planned radiation dose Distribution', respectively. The CT image used for this experiment was selected from the publicly available MIDAS database (4D CT Liver with segmentations dataset, uploaded by the Stanford School of Medicine, Stanford, California). The image was acquired on a GE Discovery ST PET/CT scanner, having a matrix and a voxel volume of $512 \times 512 \times 135$ and $0.97 \times 0.97 \times 2.5$ mm^3, respectively. The corresponding sCBCT was synthesized by simulating a reconstruction based on 120 x-Ray projections, resulting in an image quality comparable to that provided by a typical CBCT imaging system. For computational purposes, the images were re-sampled on a $256 \times 256 \times 128$ grid prior to registration. Similarly, the MR image was selected from the publicly available National Biomedical Imaging Archive (NBIA) hosted by the National Cancer Institute. The acquisition parameters, as indicated by the image meta-data, are as follows: TE = 1.06 ms, TR = 2.88 ms, flip angle = 15°, matrix size = $512 \times 512 \times 109$, voxel size = $0.78 \times 0.78 \times 2.4$ mm^3, acquired on a Siemens Avanto 1.5 T scanner. Prior to registration, the MR images were re-sampled on a $256 \times 256 \times 109$ grid, for computational purposes.

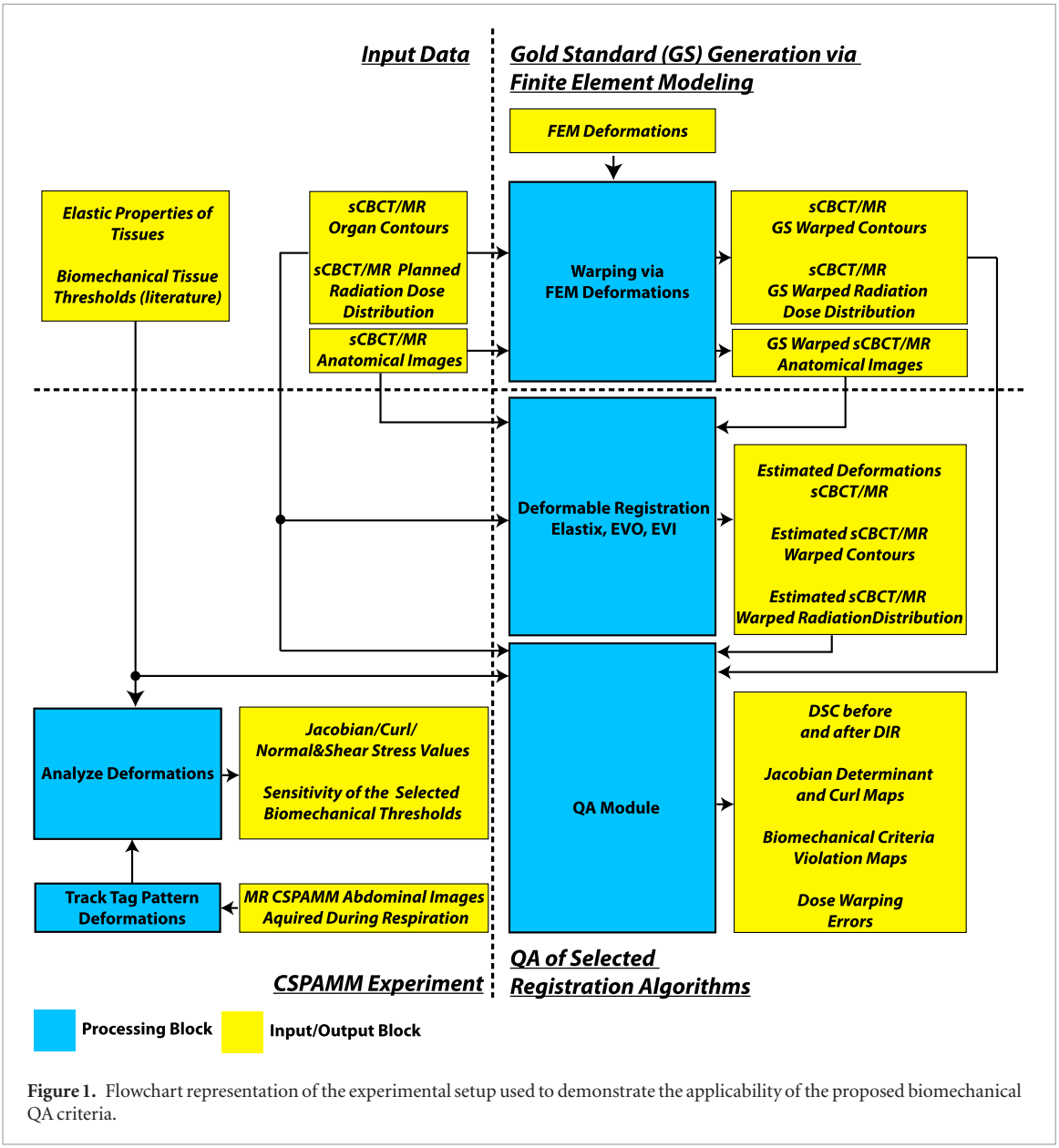

Figure 1. Flowchart representation of the experimental setup used to demonstrate the applicability of the proposed biomechanical QA criteria.

Table 1. Elastic moduli for several tissues of interest in the scope of the study.

Tissue	Elastic modulus (Pa)
Liver (Sirli *et al* 2013, Xie *et al* 2018)	6×10^3
HCC (Lu *et al* 2015)	34×10^3
Kidney (Hostettler *et al* 2010, Radulescu *et al* 2018)	20×10^3
Cancellous Bone (Lai *et al* 2015)	350×10^6

Table 2. Yield strength in tension, compression and shearing employed for the tissues of interest in the scope of this study.

Tissue	Tensile/compressive Yield strength (Pa)	Shear yield Strength (Pa)
Liver, HCC (Gao *et al* 2009, Kemper *et al* 2013)	$\pm 40 \times 10^3$	$\pm 10^3$
Kidney (Farshad *et al* 1999, Nasseri *et al* 2002)	$\pm 150 \times 10^3$	$\pm 10^3$
Trabecular bone (Kopperdahl and Keaveny 1998, Garrison *et al* 2011)	$\pm 2 \times 10^6$	$\pm 3 \times 10^6$

2.2.2. CSPAMM Experiment: determining mechanical stress limits within the human abdomen

In order to determine how the selected biomechanical thresholds reported in table 2 compare to the typical mechanical stress values occurring inside the human abdomen, an independent measurement was conducted. In this manner, the sensitivity of the selected QA-thresholds to misregistrations can be evaluated. In case the thresholds are too close to typical mechanical stress values occurring in the abdomen, any slight misregistration may violate the proposed QA-threshold and thus the probability of false positives may end-up being prohibitively high. By the same token, QA-thresholds too far beyond the values typically encountered, would lead to a very insensitive QA criteria which would be prone to false negatives.

The independent measurement of typical stress values was achieved by tracking the true deformations of the human liver and kidneys during respiration, based on a series of CSPAMM tagged MR images (Ibrahim 2011). The CSPAMM MR acquisition technique consists in the spatial modulation of the tissue spin magnetization by a set of RF saturation pulses, which appears within the acquired image as a grid of uniformly distributed low intensity areas (see figure 2). As the spins re-focus, the applied magnetization profile temporarily persists within the imaged tissues, allowing the subsequently acquired images to capture their potential deformations.

Within the current work, a series of CSPAMM images were acquired on the abdomen of five healthy volunteers. The acquisition sequence was respiratory triggered, with the tagging profile being applied at full inhalation, followed by a rapid subsequent acquisition of 10 images, sampling abdominal deformations as the volunteers exhaled. The deformations of the applied tag patterns and implicitly that of the observed tissues were extracted for each of the five volunteers via a set of tuned Gabor filters as described here (Chung *et al* 2013), resulting in a total of 45 deformation fields. In order to improve the accuracy of the measured deformations, two tag patterns were independently employed for each volunteer: a 2D grid and a 2D pair of horizontal and vertical tag lines applied separately (see figure 2). The deformations extracted from the two patterns were then averaged. The resulting deformations for the liver and one of the kidneys for all volunteers were then used to determine typical values for the Jacobian determinant, curl magnitude, tensile/compressive stress and shear stress within the organs during respiratory motion.

In terms of acquisition parameters, the CSPAMM was a gradient field echo 2D sequence employing a multi-shot echo planar imaging technique, TE $= 8.75$ ms, TR $= 90$ ms, matrix size 224×224, 25° flip angle, voxel size $1.75 \times 1.75 \times 5$ mm^3, a 5 mm distance between the tag lines and with an average acquisition time of $\sim$4 min, depending on the respiratory frequency of each volunteer. Acquisition was performed on a Philips Achieva 1.5 T MRI scanner (Philips Healthcare, Best, The Netherlands).

2.2.3. Gold standard generation using finite element modeling

For both the sCBCT and the MR image (see section 2.2.1), their corresponding contours were independently imported into a finite element modeling software (PreView v 2.1.0) (Maas *et al* 2012). Herein, a pressure of 400 Pa was applied from within the lungs, corresponding to the approximate gauge pressure within the lungs during deep inhalation. The resulting lung inflation was then simulated via the FEBio software (v 2.7.1) (Maas *et al* 2012), with the lungs acting upon the rest of the anatomical structures, simulating thoracic and abdominal deformations during inhalation. Figure 3 displays the anatomical setup extracted from the MR data and used in the finite element simulation both without and with the superimposed deformations resulting from lung inflation. Graphical rendering of the anatomical structures illustrated in figure 3, together with their displacements has been performed using the PostView software (v 2.3) (Maas *et al* 2012).

The employed elastic modulus and Poisson ratio for the liver, HCC, kidneys and ribs were the same as those reported in section 2.2.1. For the rest of the tissues, the two parameters are reported in table 3. In terms of boundary conditions, the area around the spinal column and the caudal part of the bowels were assumed to be fixated. Also, the organ interfaces were assumed to slide against one-another, except for the contacts with the adipose tissue for which a tied elastic interface was considered. Further details related to the implementation of the finite element solver employed in this study can be found in Maas *et al* (2012).

The deformations resulting from the finite element simulation (identified by the 'FEM deformations' block in figure 1) were subsequently used to map/warp the sCBCT and MR anatomical images, their associated contours and their corresponding planned radiation dose distribution, with the resulting mapped/warped data being established as gold standard for subsequent experiments.

2.2.4. QA of selected registration algorithms

The sCBCT and the 3D MR image were each registered to their FEM-deformed counterpart using a selection of three DIR solutions: the open source Elastix registration software (Klein *et al* 2010, Shamonin 2013) and two algorithms previously developed by our group. The latter consisted of the EVolution algorithm (described in detail here de Senneville *et al* (2016)) and its recently proposed incompressibility-regularized version (Zachiu *et al* 2018). From the Elastix toolbox we have selected a registration solution which maximizes the normalized mutual information (Studholme *et al* 1999) between the images to be registered and which employs a cubic

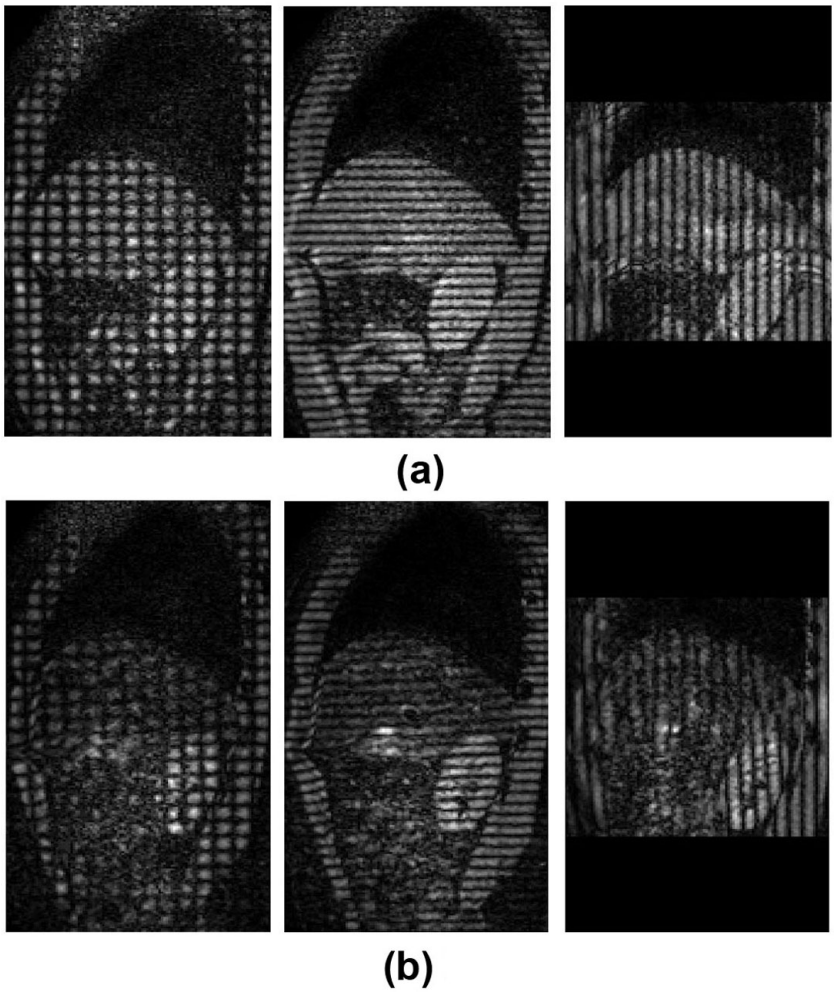

Figure 2. Example of CSPAMM tagged images showcasing a 2D grid pattern (first column) and a horizontal/vertical line pattern (second and third column). The images were acquired on the abdomen of a healthy volunteer at (a) inhalation and (b) exhalation.

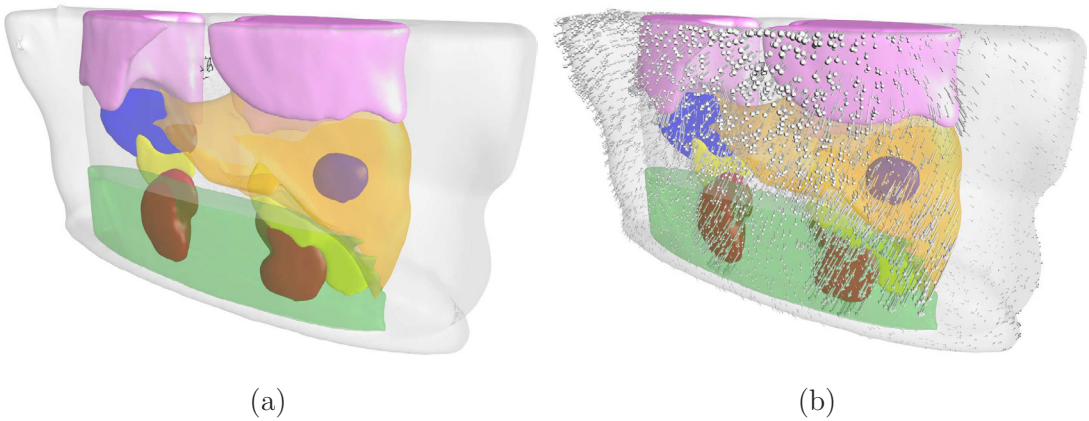

Figure 3. Anatomical setup used in the finite element simulation for the MR data set (a) without and (b) with the superimposed deformations resulting from lung inflation. The color scheme for the rendered anatomical structures is as follows: gray (body), pink (lungs), orange (liver), purple (tumor), blue (spleen), yellow (fat), red (kidneys) and green (intestinal tract). It can be seen in figure (b) that the abdominal structures undergo a predominantly caudal-anterior shift, with an additional thoraco-abdominal expansion, mainly in the ventral direction.

B-spline transformation model (Unser 1999). According to the recommendations of the toolbox authors, such a selection should be suitable for the majority of the deformable image registration tasks. On the other hand, the original EVolution algorithm (which will be abbreviated as EVO in the scope of this paper) estimates the deformation between two images I_1 and I_2 as the minimizer of the following functional:

Table 3. Elastic moduli and Poisson ratio of several tissues of interest for the finite element simulations.

Tissue	Elastic modulus (Pa)	Poisson ratio
Lung (Al-Mayah *et al* 2009)	3.74×10^3	0.4
Spleen (Arda *et al* 2011)	2.9×10^3	0.49
Bowels (Chai *et al* 2010)	3×10^3	0.35
Body (Al-Mayah *et al* 2009)	6×10^3	0.4
Fat (Comley and Fleck 2010)	1×10^3	0.49

$$E_{EVO}(\mathbf{u}) = \sum_{\vec{r}\in\Omega} e^{-C(\mathbf{u}(\vec{r}))} + \alpha\|\vec{\nabla}\mathbf{u}(\vec{r})\|_2^2 \tag{9}$$

where:

$$C(\mathbf{u}(\vec{r})) = \frac{\sum_{\vec{s}\in\Gamma} |\vec{\nabla} I_1(\vec{s}) \cdot \vec{\nabla} I_2(\vec{s} + \mathbf{u}(\vec{s}))|}{\sum_{\vec{s}\in\Gamma} \|\vec{\nabla} I_1(\vec{s})\|_2 \|\vec{\nabla} I_2(\vec{s} + \mathbf{u}(\vec{s}))\|_2} \tag{10}$$

whereas $\mathbf{u}$ is the 3D displacement, Ω is the image domain, $\vec{r}$ is the voxel position, $\vec{\nabla}$ is the spatial gradient operator, $\|\cdot\|_2$ is the Euclidean norm, α is a parameter linking the two terms of the functional and Γ is a symmetric neighborhood around $\vec{r}$. Additional details concerning the manner in which the EVO functional is minimized together with a detailed analysis of the algorithm performance can be found in the original paper proposing the algorithm (de Senneville *et al* 2016).

The incompressibility-regularized EVolution algorithm (abbreviated as EVI in the scope of this study), on the other hand, estimates the deformations between two images as the minimizer of the functional:

$$E_{EVI}(\mathbf{u}) = \sum_{\vec{r}\in\Omega} e^{-C(\mathbf{u}(\vec{r}))} + \beta\|J(\vec{r} + \mathbf{u}(\vec{r})) - 1\|_2^2 \tag{11}$$

where J is the Jacobian determinant of the estimated deformation. All other terms preserve their meaning from equation (9). Further details related to EVI can be found in Zachiu *et al* (2018).

The deformations estimated by the three registration algorithms were tested against both the proposed biomechanical QA criteria and the existing Jacobian determinant and curl magnitude criteria. Note that, in the scope of this experiment, all Jacobian, curl and mechanical stress calculations were performed in 3D. The estimated deformations were also used to map/warp the contours associated to the sCBCT and the MR image and their respective planned dose distributions. This allowed the evaluation of the contour alignment and dose mapping accuracy of the three methods with respect to the FEM-generated gold standard. More importantly, however, in particular for dose mapping, this has allowed to analyze the spatial correspondence between violations of the proposed biomechanical QA criteria and potential dose mapping errors.

2.3. Implementation and configuration of the registration algorithms

Both EVO and EVI were implemented using the compute unified device architecture (CUDA) framework and executed on an nVidia Tesla K20 graphical processing unit (GPU). The values of the control parameters α, β and Γ (see equations (9) and (11)) was maintained the same as in Zachiu *et al* (2018). For configuring Elastix, we have followed the recommendations provided within the utilization manual by the toolbox authors. The deformation vector fields estimated by each of the algorithms were then provided as an input to a multi-threaded (8 threads) C++ QA module implementing the criteria proposed in the current study.

3. Results

3.1. Typical deformations of the liver and kidneys

Figure 4 illustrates the statistical distribution of the Jacobian determinant and curl magnitude of the liver and kidney deformations determined by tracking the CSPAMM tag patterns during respiration (see section 2.2.2 for details). The boxplots have been generated by pooling together the two parameters from all five of the healthy volunteers. This resulted in the Jacobian determinant of the deformations being within the [0.85 1.1] range for the liver and within [0.94 1.07] for the kidney. Concerning the curl magnitude of the deformations, this attained a maximum of 0.2 and 0.1 for the liver and kidney, respectively.

Figure 5 displays the range of the tensile/compressive and shear stress within the human liver and kidneys during respiration, determined by pooling together the values obtained from all five of the volunteers. Thus, it was observed that the tensile/compressive stress is within [−720212 016] Pa for the liver and [−16 71514 200] Pa for the kidney, while the shear stress is within [−14993] Pa for the liver and [−184224] Pa for the kidney.

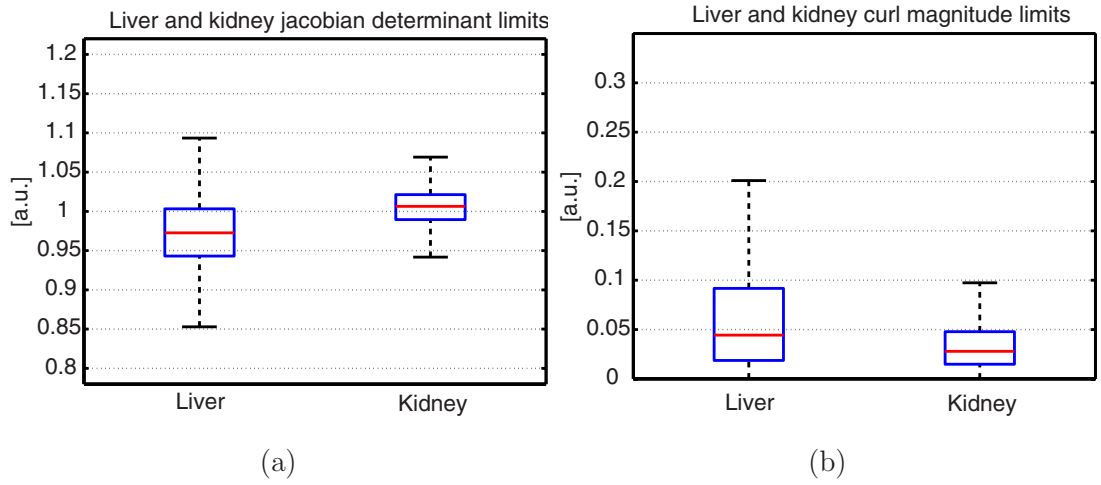

Figure 4. Statistical distribution of the (a) Jacobian determinant and (b) curl magnitude for the liver and kidney deformations, determined by tracking CSPAMM tag patterns during respiration.

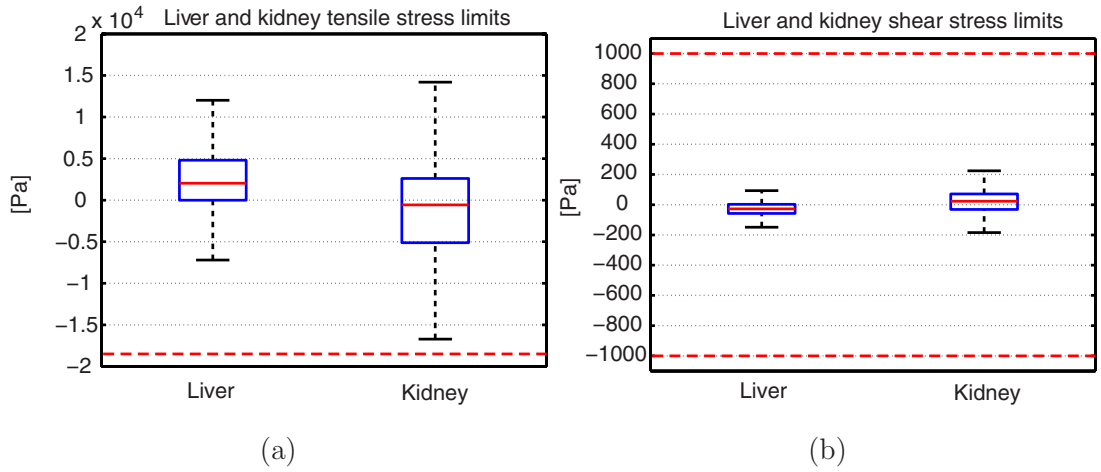

Figure 5. Statistical distribution of the (a) tensile/compressive and (b) shear mechanical stress within the liver and kidney during respiration, as determined by tracking CSPAMM tag patterns. The red dashed line in figure (a) indicates the compressive stress at which arterial blood vessel occlusion would start occurring, while in figure (b) it indicates the shear stress required for tissue plastic deformation. Note that these limits were adopted directly from section 2.2.1 and were not determined in the scope of the CSPAMM experiment.

3.2. Quality assurance for CBCT-based therapy guidance

Table 4 reports the pre- and post-registration DSC for several volumes-of-interest (VOI) within the sCBCT dataset. It can be observed that all of the employed registration algorithms led to a notable improvement of the DSC for all VOIs, with DSC $>$ 0.9 in all cases. Comparatively, the three algorithms provide similar DSC values with marginal differences of 1%–2%.

Figure 6(a) showcases a sagittal slice from the sCBCT used as reference during the registration process of the two sCBCT images (see section 2.2.3). This is followed in figures 6(b)–(g) by the Jacobian determinant and the curl magnitude within the same slice of the deformations estimated by the three employed registration algorithms for the VOIs indicated in table 4. It can be observed that the three algorithms showcase varying amounts of fluctuations of the Jacobian determinant from one. The deformations estimated by both Elastix and EVO manifest moderate deviations from unity of the Jacobian determinant with isolated peaks within both the liver and the kidney. By comparison, EVI demonstrates considerably lower deviations of the Jacobian determinant from one. Similar observations can be made with respect to the curl magnitude, with all algorithms predominately showcasing low to moderate values of 0–0.4.

Figure 7 showcases the voxels in which the three proposed biomechanical QA criteria have detected physically implausible deformations, with respect to the thresholds provided in section 2.2.1. The illustration is performed as an overlay on the sagittal slice displayed in figure 6(a), for all three of the employed registration algorithms. The voxels in which the estimated deformations lead to a surplus of the tensile/compressive stress for plastic deformations are indicated in figures 7(a)–(c) as red and blue overlays, respectively. It can be observed that while

Table 4. Contour alignment capabilities of the investigated registration algorithms. The table reports the pre- and post-registration DSC attained by each registration method for several volumes of interest.

	Algorithm			
Volume of interest	None	Elastix	EVO	EVI
GTV	0.69	0.92	0.91	0.92
Liver	0.91	0.97	0.97	0.98
Kidneys	0.89	0.97	0.97	0.98
Bones	0.85	0.95	0.95	0.97

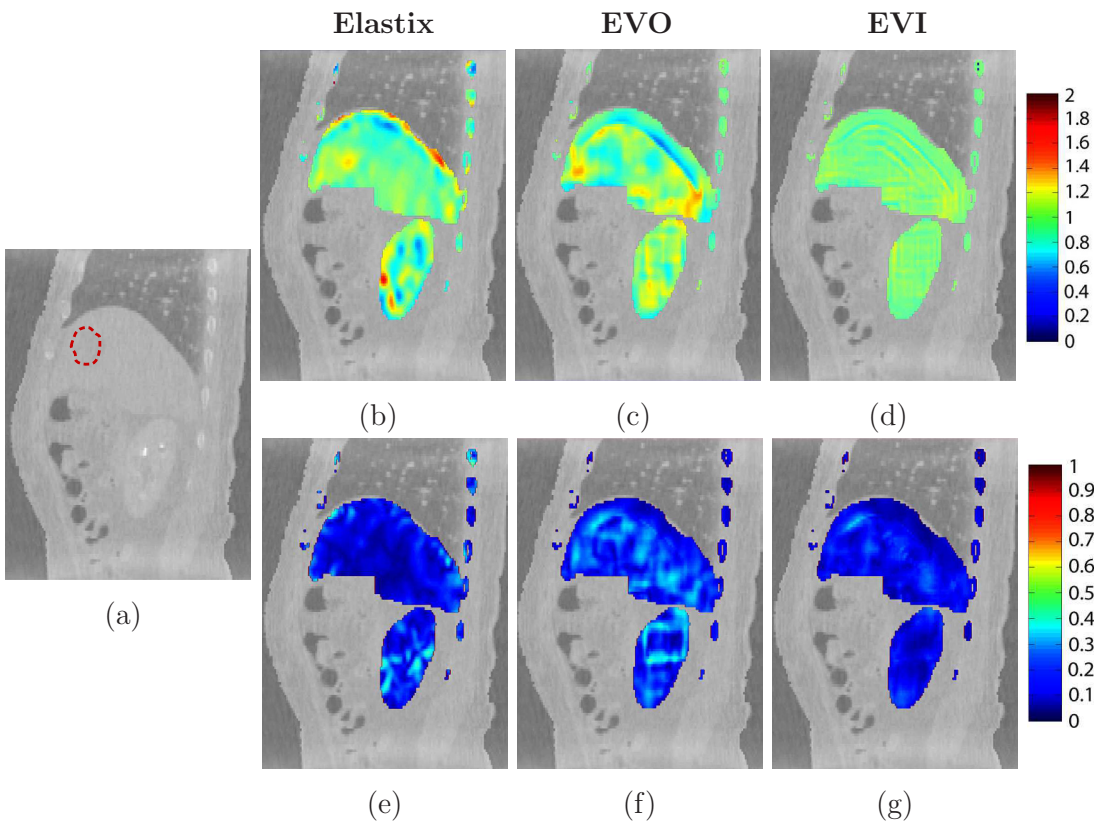

Figure 6. Anatomy and differential quality assurance criteria for the sCBCT registration. (a) A sagittal slice selected from the reference sCBCT image, originally acquired on a hepatocellular carcinoma patient. The red dashed line indicates the location of the GTV. (b)–(d) The Jacobian and (e)–(g) the curl magnitude of the deformations estimated by the three registration algorithms within the sagittal slice illustrated (a).

Elastix showcases moderate violations of this criterion within the GTV, such problematic areas become more extensive for EVO. In addition, the Elastix also leads to extensive implausible deformations within the ribs. EVI, on the other hand, considerably reduces the extent of these violations, with only the GTV showcasing a few islets of anatomically implausible deformations with respect to this criterion.

In terms of the shear stress exceeding physiological limits, as illustrated in figures 7(d)–(f), Elastix showcases a moderate amount of problematic areas within the ribs and the kidney, while EVO to a similar extent within the GTV and the kidney. EVI on the other hand reduces this extent to overall a few isolated voxels (not visible on the displayed slice). The areas in which the estimated deformations lead to a compressive pressure exceeding that of the typical arterial blood pressure are showcased as a red overlay in figures 7(g)–(i). It can be observed that such violations occur over extensive regions for all algorithms. While the deformations estimated by EVI again seem to have a higher plausibility compared to Elastix and EVO, rather wide areas within the GTV and the kidney still remain problematic. Also note that there is good spatial correspondence between the areas 'flagged-up' in figures 7(g)–(i) and the areas with a low Jacobian determinant in figure 6.

3.3. Quality assurance for MR-based therapy guidance

A similar analysis as in the previous section was also performed for the deformations estimated by Elastix, EVO and EVI on the pair of 3D MR images (see section 2.2.3). Table 5 showcases the contour alignment capabilities of the three algorithms for the MR image pair. It can be observed that for all volumes-of-interest, the algorithms

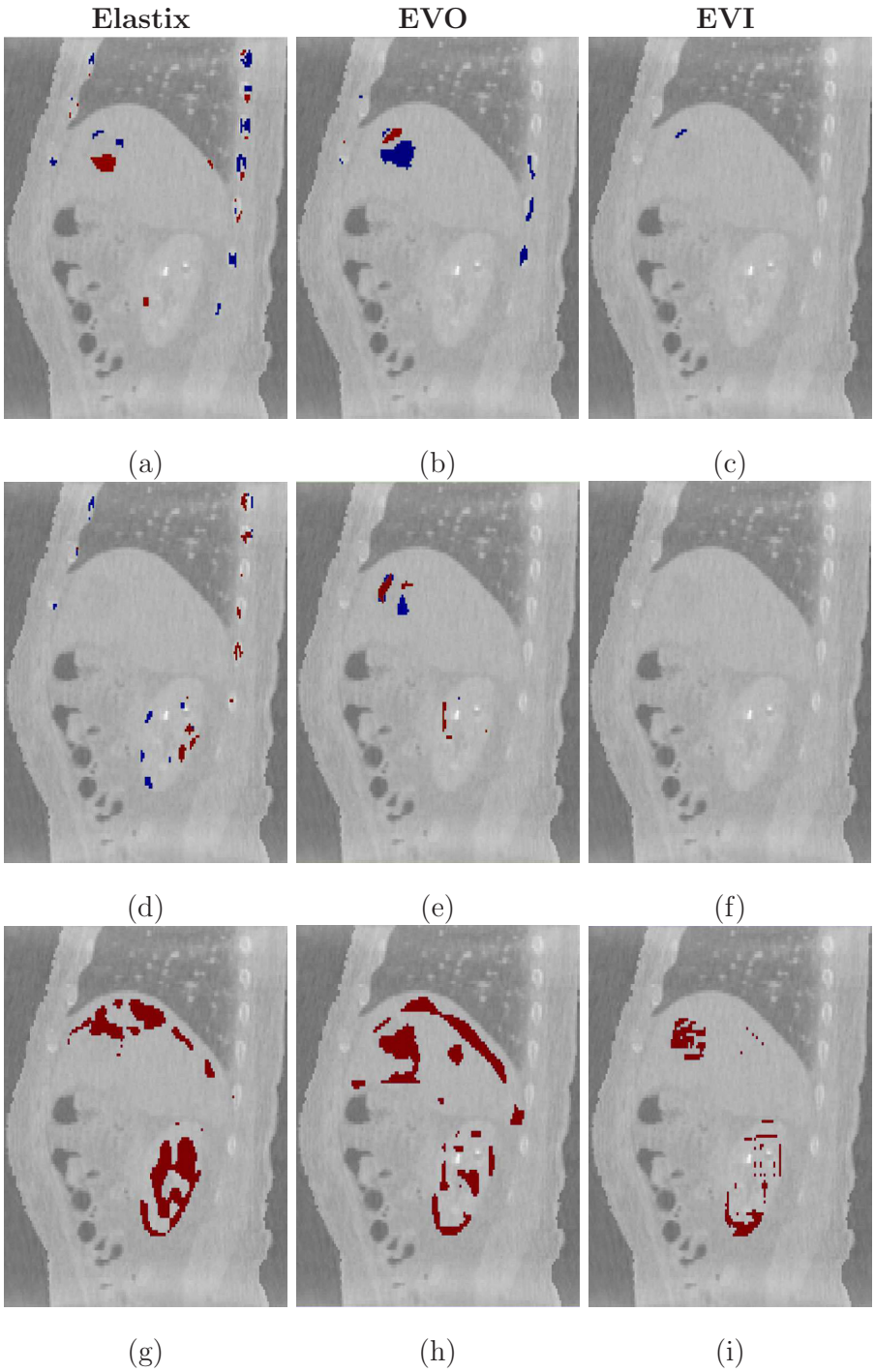

Figure 7. Bio-mechanical quality-assurance criteria for the sCBCT registration. (a)–(c) Voxels in which the estimated deformations indicate an surplus of the tensile (in red) and compressive (in blue) stress for plastic deformations, as given in table 2. (d)–(f) Voxels in which the estimated deformations indicate an excess of the shear stress for plastic deformations (see table 2). The blue and the red colors indicate the sign of the shear (negative and positive, respectively). (g)–(i) Voxels in which the estimated compressive deformations indicate an excess of the typical arterial blood pressure.

led to an improvement of the DSC. Similar to the case of sCBCT (see table 4), the three methods demonstrate a similar capability for contour alignment, with 1%–2% differences in terms of the DSC.

Figure 8(a) displays sagittal slice from the MR image used as reference during the registration process, with the GTV appearing as an abnormality within the posterior part of the liver. This is followed in figures 8(b)–(g) by the spatial distribution within the same slice of the Jacobian determinant and the curl magnitude of the deformations estimated within the liver and one of the kidneys by the three employed registration algorithms. Both Elastix and EVO showcase overall moderate deviations of the Jacobian determinant from unity, with isolated large peaks. EVI, on the other hand, provides deformations with considerably lower deviations of the Jaco-

Table 5. Contour alignment capabilities of the investigated registration algorithms, for the 3D MR image pair. The table reports the pre- and post-registration DSC attained by each registration method for several volumes of interest.

	Algorithm			
Volume of interest	None	Elastix	EVO	EVI
GTV	0.84	0.94	0.94	0.96
Liver	0.94	0.97	0.97	0.98
Kidneys	0.89	0.95	0.95	0.96

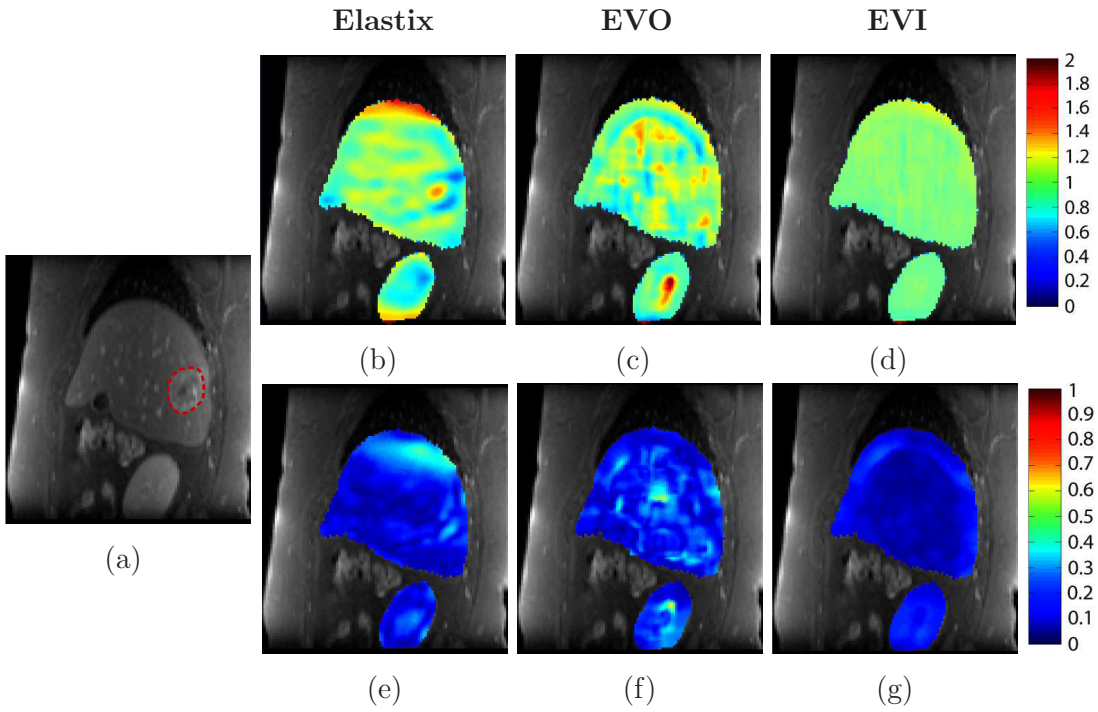

Figure 8. Anatomy and differential quality assurance criteria for the MR registration. (a) A sagittal slice selected from the reference MR image, acquired on a hepatocellular carcinoma patient. The red dashed line indicates the location of the GTV. (b)–(d) The Jacobian and (e)–(g) the curl magnitude of the deformations estimated by the three registration algorithms within the sagittal slice illustrated in (a).

bian determinant from one. As for the curl magnitude, all algorithms provide predominantly low to moderate values, with sparse moderate peaks for the Elastix and EVO algorithms.

Figure 9 illustrates the anatomical plausibility of the deformations estimated by the three algorithms on the 3D MR image pair, according to the proposed biomechanical QA criteria. The illustration was limited to the sagittal slice from figure 8(a). With respect to the tensile/compressive stress limits for plastic deformations (figures 9(a)–(c)), both Elastix and EVO showcase large problematic areas within the GTV, while at the same time leading to more isolated regions within the liver and the kidney. EVI, on the other hand, did not violate the criterion in any of the voxels. A good correspondence can be observed for Elastix and EVO between the areas flagged by the tensile/compressive criterion and the regions with high deviations of the Jacobian determinant from unity in figure 8.

For the shear stress criterion (figures 9(d)–(f)), the Elastix and the EVO algorithms showcase several problematic areas within the GTV and the kidney. Note that there is an observable spatial correspondence between the shear stress flags from figure 9(e) and the peaks within curl magnitude map from figure 8(f). The deformations estimated by EVI did not raise any flags with respect to the yield shear stress criterion.

As it can be noted from figures 9(g)–(h), both the Elastix and the EVO algorithms provided deformations which violate the blood pressure criterion rather extensively, within all the VOIs. The EVI algorithm leads to a limited amount of violations within the kidney and the GTV, with respect to this particular criterion.

3.4. Relevance of the proposed biomechanical QA criteria for dose warping

The deformations estimated by the three registration algorithms on both the sCBCT and the MR datasets were used for dose mapping in the manner described in section 2.2.4. The deformations resulting from the finite

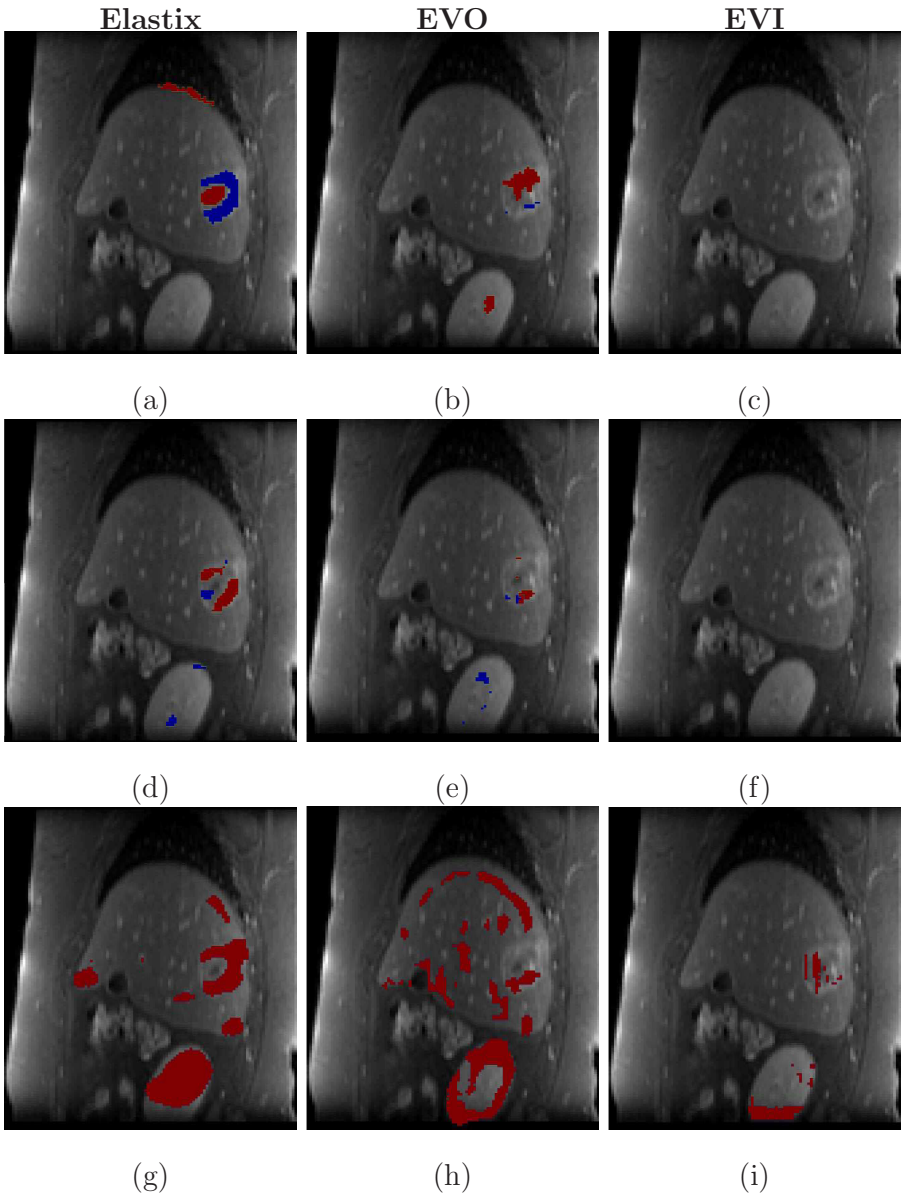

Figure 9. Proposed bio-mechanical quality-assurance criteria. (a)-(c) Voxels in which the estimated deformations indicate an excess of the tensile (in red) and compressive (in blue) stress for plastic deformations given in table 2. (d)–(f) Voxels in which the estimated deformations indicate an excess of the shear stress for plastic deformations (see table 2). The red and the blue colors indicate the sign of the shear (positive and negative, respectively). (g)–(i) Voxels in which the estimated compressive deformations indicate a surplus of the typical arterial blood pressure.

element simulation fulfilled all the proposed biomechanical QA criteria, with no violations being detected within either of the soft tissues of interest. Furthermore, the Jacobian determinant, curl magnitude and the mechanical stress values corresponding to these deformations were within the same range as the ones provided by the CSPAMM experiment.

Figure 10(a) displays for the sCBCT data, the dose mapped using the deformations provided by the three algorithms as well as the deformations resulting from the finite element simulation (used as gold standard). It can be observed that compared to the gold standard, the three mapped dose distributions showcase a varying amount of differences, with error peaks spatially correlated with violations of the proposed QA criteria as illustrated in figure 7. From a visual point-of-view, the dose mapped by the EVI algorithm showcases the best similarity to the gold standard, followed by Elastix and lastly EVO.

A more objective comparison regarding the anatomical plausibility of the deformations provided by the three algorithms, together with its impact on the mapped dose is displayed by figures 10(b)–(d). Figure 10(b) displays the percentage of GTV voxels flagged by each of the biomechanical QA criteria applied for the deformations estimated by each of the three algorithms. As already anticipated from figure 7, Elastix and EVO showcase a notably larger amount of voxels in which the tensile/compressive and the shear stress criteria are violated, compared to EVI. However, in terms of the blood pressure criterion, EVI exceeds Elastix. Figure 10(c), on the

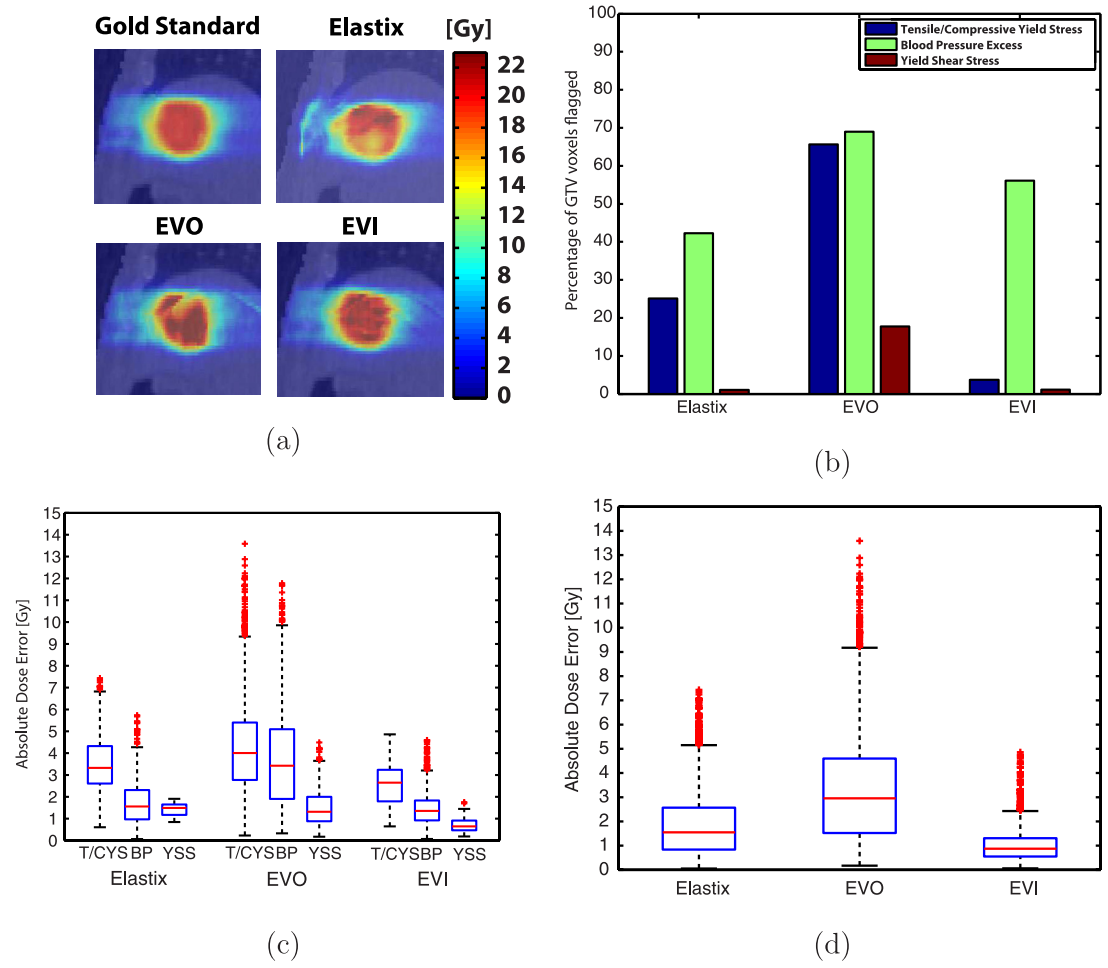

Figure 10. Results of the dose mapping experiment conducted on the sCBCT image pair, showcasing the correspondence between violations of the proposed biomechanical QA criteria and errors within the mapped dose distribution. (a) Mapped dose distribution using the deformations provided by the finite element simulation and the three registration algorithms. (b) Percentage of the GTV voxels 'flagged' by the proposed biomechanical QA criteria for each of the three employed algorithms. (c) Absolute dose warping errors within the voxels 'flagged' by each of the three biomechanical quality assurance criteria: tensile/compressive yield stress (T/CYS), blood pressure (BP) excess and yield shear stress (YSS). (d) Absolute dose mapping errors within the entire GTV.

other hand, illustrates the relationship between violations of the proposed biomechanical QA criteria within the GTV and the absolute errors in the mapped dose distribution. Notice that violations of the tensile/compressive stress criterion appear to be the source of the highest dose mapping inaccuracies for all algorithms, with an error median of ∼3–4 Gy and a maximum of ∼14 Gy (∼70% of the maximum dose) attained for the EVO registration method. The blood pressure excess, while leading on average to lower dose mapping errors, has a similar dose error range as the tensile/compressive stress criterion for each individual algorithm. Shear stress violations, on the other hand, have led to significantly lower errors compared to both the blood pressure and the tensile/compressive stress criteria. Figure 10(d) displays the absolute dose mapping errors pooled from the entirety of the GTV. Of all algorithms, EVO leads to the highest errors, followed by Elastix and EVI.

Figure 11 provides a similar analysis as figure 10 for the MR dataset. The three algorithms again showcase varying degrees of discrepancy in the mapped dose compared to the gold standard, however, in this case it can be observed that Elastix provides the highest errors, followed by EVO and EVI. For the latter, the mapped dose distribution is, from a visual perspective, borderline indiscernible from the gold standard. These observations are further confirmed by figures 11(b)–(d), where it can be noted that Elastix violates the tensile/compressive stress criterion in a considerably larger fraction of voxels compared to EVO and EVI. Additionally, while the EVO algorithm provides similar results as in the sCBCT case, note that EVI no longer violates the tensile/compressive and the shear stress criteria. It is only the blood pressure criterion which still remains to different extents problematic for all three algorithms. The statistical analysis of the mapped dose errors displayed in figures 11(c)–(d) reveals a notably lower error range for EVO and EVI compared to the sCBCT case. For Elastix on the other hand, there is a considerable increase of the dose errors overall, as also illustrated in figures 11(a) and (c), with EVI, as previously observed, demonstrating the lowest dose error range.

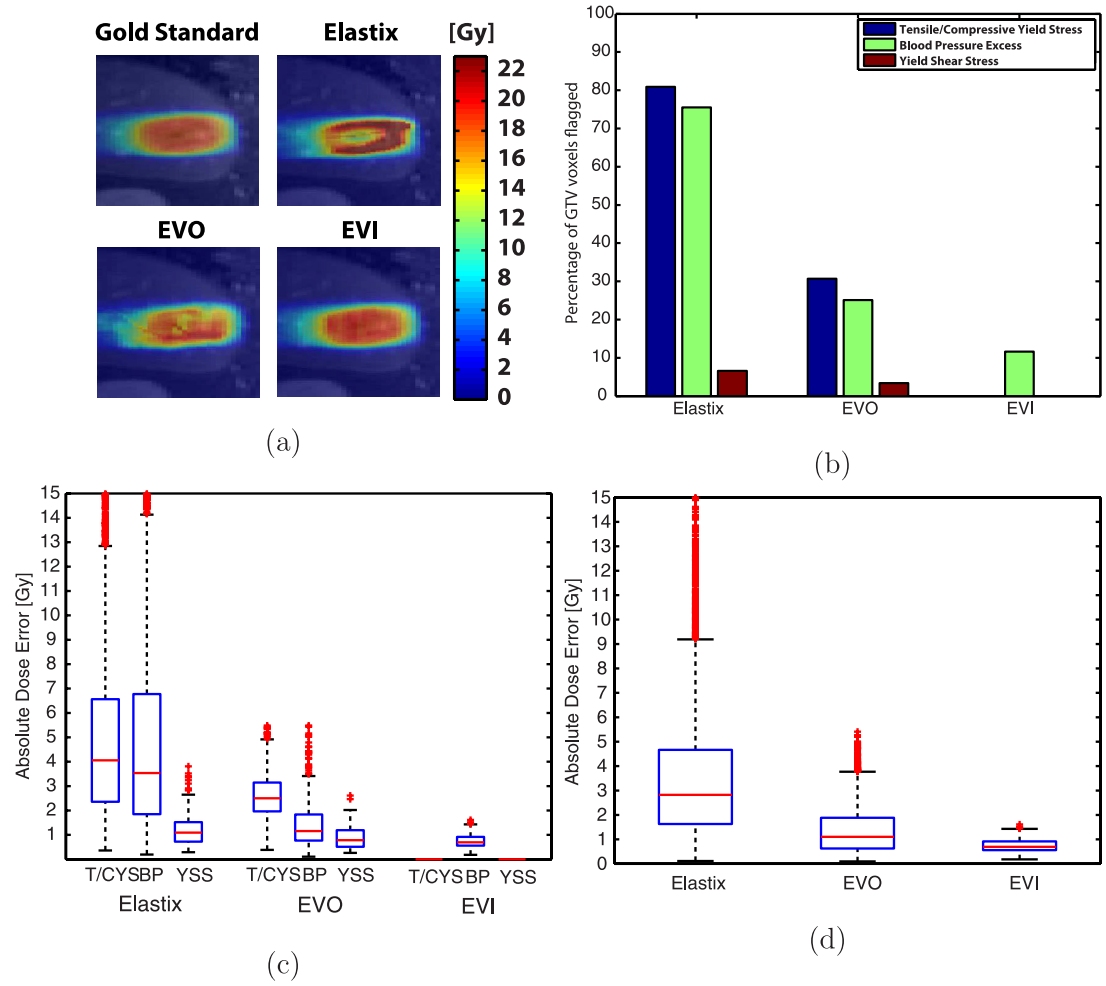

Figure 11. Results of the dose mapping experiment conducted on the 3D MR image pair, showcasing the correspondence between violations of the proposed biomechanical QA criteria and errors within the mapped dose distribution. (a) Mapped dose distribution using the deformations provided by the finite element simulation and the three algorithms. (b) Percentage of the GTV voxels 'flagged' by the proposed biomechanical quality assurance criteria for each of the three employed algorithms. (c) Absolute dose mapping errors within the voxels 'flagged' by each of the three biomechanical quality assurance criteria: tensile/compressive stress for plastic deformations (T/CYS), blood pressure (BP) excess and shear stress for plastic deformations (YSS). (d) Absolute dose mapping errors within the entire GTV.

4. Discussion

Multiple concepts from adaptive IGRT such as day-to-day repositioning, daily re-planing, contour propagation and dose accumulation rely on the deformations provided by mono- and multi-modal registration algorithms (Brock *et al* 2017). Therefore, in order to ensure patient safety, the estimated deformations used for IGRT adaptation should preferably be subjected to a battery of complementary QA criteria. However, as emphasized over the course of the present study, the selection of such criteria is not a trivial task, particularly when they are required to operate in online applications and thus automated.

4.1. Quality assurance potential of existing criteria

Current clinical practices for validation of medical image registration results in an online setting is visual inspection of high contrast anatomical landmarks such as organ boundary alignment. In the scope of this work the Dice similarity coefficient (DSC), which measures organ contour overlap, was selected as an objective surrogate for a visual analysis of organ boundary alignment. As indicated by tables 4 and 5, all three of the analyzed registration algorithms provide DSC values within 1%–2% of each other and therefore demonstrating similar contour alignment capabilities for the two analyzed datasets. However, as emphasized before, a high-fidelity alignment of the organ boundaries does not guarantee that within the boundaries the estimated deformations are accurate. Consequently, as suggested by previous studies, the Jacobian determinant and the curl of the estimated deformations have been investigated within the organ boundaries.

Both Elastix and EVO have showcased moderate deviations of the Jacobian determinant from one, with a number of locally high/low peaks, in particular for the 3D MR dataset (see figure 8). By comparison, these devia-

tions are considerably lower for the EVI algorithm, indicating a better correspondence with the physical reality. Such an outcome is, however, simply a proof of convergence for the EVI algorithm, since deviations of the Jacobian determinant from unity are explicitly penalized by the employed registration model (see equation (11). The occurrence of vortices is, however, not penalized and will appear as peaks within the curl magnitude map and therefore detectable. It was observed, nevertheless, that deep within the organ boundaries, the EVI algorithm showcases the lowest overall values of the curl magnitude of all three of the analyzed registration algorithms (see figures 6 and 8). Comparing the Jacobian determinant and curl magnitude of deformations within soft tissues to one and zero, respectively, is, however, of borderline qualitative value for QA. Depending on their mechanical properties, biological tissues may allow a particular degree of compression/expansion or torsion (as demonstrated in figure 4). Therefore, without taking into account these mechanical properties, it may be challenging to decide whether a particular value for the Jacobian determinant or curl magnitude is anatomically plausible or not. The proposed biomechanical QA criteria come to address this issue and to a particular degree relate the Jacobian determinant and curl magnitude to measurable physical quantities.

4.2. Quality assurance potential of the proposed biomechanical criteria

Among all three criteria, T/CYS criterion detects the most anatomically implausible deformations, indicated not only by the high thresholds on the resulting stress/pressure (see table 2) but also by corresponding to the highest errors in terms of the mapped radiation dose (see figures 10 and 11). It was also observed that is that there is a good correspondence between violations of the T/CYS criterion and strong deviations of the Jacobian determinant from one (see figures 8 and 9). In effect, the T/CYS criterion can provide a tissue-dependent interpretation of the acceptable value range of the Jacobian determinant.

The proposed blood pressure excess (BP) criterion can be seen as a 'superset' of the compressive stress for plastic deformations criterion. This is due to the fact that if the T/CYS indicates a compressive violation, the BP criterion will also respond. This may explain the reason for which the two criteria correspond to a similar dose mapping error range (see figures 10(c) and 11(c)), since errors corresponding to compressive T/CYS violations are also included by BP violations. Out of all three criteria, BP has demonstrated the highest sensitivity, indicated by the extensive violations it detects for all three of the analyzed registration algorithms. This may prove to be beneficial in detecting more minor potential errors within the mapped dose distribution. While for linear dose accumulation (i.e. dose mapping, without feeding the obtained/delivered dose into a new plan) such errors may have a clinically/therapeutically acceptable impact, for adaptive strategies which iteratively feed the mapped dose into the therapeutic dose plan, the errors may be potentially amplified over feedback.

Similar to how the T/CYS and the BP criteria can provide a tissue-dependent interpretation of the Jacobian determinant, the YSS criterion can be seen as a tissue-dependent interpretation of the curl magnitude of the deformations. Strong local variations within curl magnitude maps are indicative of the occurrence of shearing motion. The YSS criterion indicates whether this shearing motion is anatomically plausible or not, depending on the mechanical properties of the observed tissues. This correspondence can be observed by comparing figures 6 and 8 to figures 7 and 9, respectively, where strong spatial variations within the curl magnitude maps correspond to violations of the YSS criterion. It was also observed that, among all three criteria, the YSS provided considerably lower dose mapping errors by comparison (see figures 10(c) and 11(c)).

An intrinsic limitation of the proposed criteria is that therapeutic response may potentially be flagged-up as an image registration failure. Developing generally-applicable QA criteria for deformations estimated in such a case is particularly challenging, since over the course of the treatment the tumor may undergo a broad range of volumetric and radiographic changes, in addition to a potential weight loss of the patient as a side-effect. Thus, for inter-fraction adaptation, the proposed biomechanical QA criteria should be interpreted with consideration towards the amount of radiological and anatomical changes underwent by the observed anatomy. Other limitations include the applicability to anatomical structures such as the bladder, the lung and the rectum, which display large volume fluctuations over time and can thus intrinsically not be evaluated by the presented criteria.

4.3. Biomechanical modeling of the observed tissues

The elastic modulus, Poisson ratio and the tensile/compressive/shear yield strength of biological tissues has been historically determined via pure mechanical tests, with the former two also being obtainable more recent by MR and US elastography.

An alternate way to measure typical values of the mechanical strains which occur within soft tissues during physiological motion is by the means of CSPAMM MR-tagging, as demonstrated in the current work. This revealed that while the Jacobian determinant and curl magnitude of the deformations is close to one and zero, respectively, for both the liver and the kidneys, the range of the two criteria is larger for the liver (see figures 4 and 5). Particularly for compressions/expansions, this may be explained by the liver's higher content of venous blood, which when displaced it provides a mean for volume changes. Additionally, this may also be explained

by the different elastic properties of the two organs, with the kidney being stiffer than the liver, thus allowing a smaller range of deformations. Measurements of the tensile, compressive and shear stresses within the liver and kidney (see figure 5), revealed a considerably lower range compared to the selected limits reported in table 2 illustrating thus, the sensitivity of the proposed biomechanical criteria towards detecting misregistrations. In terms of the T/CYS criterion, the measured tensile and compressive stress values for the liver and kidneys are below the violation threshold by a factor of approximately four and ten, respectively (see figure 5 and table 2). For the YSS criterion, the measured shear stress for the liver and kidneys was smaller than the violation threshold by a factor of approximately six and five, respectively. Thus, we can conclude that within voxels which have been 'flagged-up' by the T/CYS and YSS criteria, there is a high probability that misregistrations have indeed occurred. For the BP criterion, on the other hand, the measured compressive stress within the liver and kidneys was smaller than the selected threshold by a factor of approximately 2.5 and 1.1, respectively. This reconfirms the higher sensitivity of the BP criterion, compared to T/CYS and YSS, having at the same time an increased probability of detecting false positives.

Within this study, all tissues under observation were assumed to have linear elastic behavior with isotropic properties. While this is a simplified model, the selected biomechanical QA criteria provide a sufficient margin compared to typical stress limits which occur during physiological motion (as previously argued), thus we do not expect a more complex model to lead to significantly different observations over the course of the conducted experiments. Nevertheless, this may constitute the object of future studies.

4.4. Evaluation of the selected registration algorithms with respect to the investigated QA criteria

Three different solutions were employed for the registration of a pair of synthetic CBCT and MR images: Elastix, EVolution (EVO) and EVolution Incompressible (EVI). We would like to underline that the purpose of the comparative analysis among the three approaches was not to determine the best performing one, but rather to showcase the response of the proposed QA criteria to deformations estimated by algorithms with different functioning principles. While all three algorithms provided similar DSC values (see tables 4 and 5), differences were observed in terms of Jacobian determinant, curl magnitude and number of violations of the proposed biomechanical criteria. Of the three methods, EVI demonstrated lower deviations of the Jacobian determinant from one and a lower number of violations of the T/CYS and BP criteria across the imaged volume. However, this can be expected since the EVI algorithm penalizes by design compressive/expansive deformations. Since EVI uses part of the criteria in the regularization, the evaluation with respect to the QA criteria does not amount to an independent evaluation, but is indirectly merely proof-of-convergence for the algorithm. In addition, while within soft tissues EVI comes across as superior to the other two methods, within anatomical areas which do showcase volumetric changes (such as for example the lung), it will most likely lead to misregistrations.

Another aspect to note is that the control parameters for the EVO and EVI algorithms were adopted directly from Zachiu *et al* (2018), while for Elastix, the recommendations of the toolbox authors were used for calibration (see section 2.3). Within Zachiu *et al* (2018), both EVO and EVI were calibrated such that the post-registration organ boundary alignment is maximized. This was meant to be a surrogate for the manner in which registration validation is currently performed in a clinical interventional setting, by ensuring that organ boundaries are properly aligned after registration. It may be possible, however, to optimize the algorithm calibration for minimizing violations of the proposed QA criteria. This may prove nevertheless to be to the detriment of other criteria. For example, if EVO was to be re-tuned for the aforementioned purpose, by increasing the value of α (see equation (9)), this will lead to a degradation of its organ boundary alignment capabilities, since the overall elasticity of the estimated deformations will decrease. Also, once an algorithm is calibrated with respect to a particular criterion, the criterion may no longer be representative for the overall QA process and thus alternatives may have to be considered.

5. Conclusion

The current study proposes a set of biomechanical criteria which come to extend the existing canon of QA for DIR of biological soft tissues. The criteria make use of the specific biomechanical properties of the observed anatomy, providing a fast and objective physical interpretation of the estimated deformations, suitable for applications with a low-latency requirement such as adaptive IGRT.

In terms of clinical value, the criteria are particularly important for adaptive workflows employing deformable registration algorithms, being indicators for potentially severe errors within the mapping process of quantitative information such as radiation dose distributions. Thus, the proposed QA solution has great potential in ensuring patient safety not only for adaptive IGRT, but also for other use-cases which require the spatial re-mapping of physical and/or physiological properties, such as Houndsfield units, contrast uptake, diffusion or perfusion values.

Acknowledgments

This work was supported by ITEA 3, project No. 16016 (STARLIT).

ORCID iDs

Cornel Zachiu https://orcid.org/0000-0001-9755-4584

References

Al-Mayah A, Moseley J, Velec M and Brock K K 2009 Sliding characteristic and material compressibility of human lung: parametric study and verification *Med. Phys.* **36** 4625–33
Arda K, Ciledag N, Aktas E, Arıbas B K and Kse K 2011 Quantitative assessment of normal soft-tissue elasticity using shear-wave ultrasound elastography *Am. J. Roentgenol.* **197** 532–6
Biguri A, Dosanjh M, Hancock S and Soleimani M 2016 TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction *BioMed. Phys. Eng. Express* **2** 055010
Bissonnette J P, Balter P A, Dong L, Langen K M, Lovelock D M, Miften M, Moseley D J, Pouliot J, Sonke J J and Yoo S 2012 Quality assurance for image-guided radiation therapy utilizing CT-based technologies: a report of the AAPM TG-179 *Med. Phys.* **39** 1946–63
Bistoquet A, Oshinski J and Skrinjar O 2008 Myocardial deformation recovery from cine MRI using a nearly incompressible biventricular model *Med. Image Anal.* **12** 69–85
Bortfeld T, Jiang S and Rietzel E 2004 Effects of motion on the total dose distribution *Semin. Radiat. Oncol.* **14** 41–51
Brock K K, Mutic S, McNutt T R, Li H and Kessler M L 2017 Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM radiation therapy committee task group no 132 *Med. Phys.* **44** e43–76
Chai X, van Herk M, van de Kamer J B, Hulshof M C C M, Remeijer P, Lotz H T and Bel A 2010 Finite element based bladder modeling for image-guided radiotherapy of bladder cancer *Med. Phys.* **38** 142–50
Chalana V and Kim Y 1997 A methodology for evaluation of boundary detection algorithms on medical images *IEEE Trans. Med. Imaging* **16** 642–52
Chung S, Kim K E, Park M S, Bhagavatula S, Babb J and Axel L 2013 Liver stiffness assessment with tagged MRI of cardiac-induced liver motion in cirrhosis patients *J. Magn. Reson. Imaging* **39** 1301–7
Comley K and Fleck N A 2010 A micromechanical model for the Young's modulus of adipose tissue *Int. J. Solids Struct.* **47** 2982–90
de Senneville B D, Zachiu C, Ries M and Moonen C 2016 EVolution: an edge-based variational method for non-rigid multi-modal image registration *Phys. Med. Biol.* **61** 7377–96
Dice L R 1945 Measures of the amount of ecologic association between species *Ecology* **26** 297–302
Doyley M M and Parker K J 2014 Elastography: general principles and clincial applications *Ultrasound Clinics* **9** 1–11
Farshad M, Barbezat M, Fleler P, Schmidlin F, Graber P and Niederer P 1999 Material characterization of the pig kidney in relation with the biomechanical analysis of renal trauma *J. Biomech.* **32** 417–25
Feldkamp L A, Davis L C and Kress J W 1984 Practical cone-beam algorithm *J. Opt. Soc. Am.* A **1** 612
Gao Z, Lister K and Desai J P 2009 Constitutive modeling of liver tissue: experiment and theory *Ann. Biomed. Eng.* **38** 505–16
Garrison J G, Gargac J A and Niebur G L 2011 Shear strength and toughness of trabecular bone are more sensitive to density than damage *J. Biomech.* **44** 2747–54
Guckenberger M 2011 Image-guided radiotherapy based on kilovoltage cone-beam computed tomography—a review of technology and clinical outcome *Eur. Oncol. Haematol.* **07** 121
Hostettler A, George D, Rémond Y, Nicolau S A, Soler L and Marescaux J 2010 Bulk modulus and volume variation measurement of the liver and the kidneys *in vivo* using abdominal kinetics during free breathing *Comput. Methods Programs Biomed.* **100** 149–57
Ibrahim E S H 2011 Myocardial tagging by cardiovascular magnetic resonance: evolution of techniques–pulse sequences, analysis algorithms, and applications *J. Cardiovascular Magn. Reson.* **13** 36
Jaffray D A 2012 Image-guided radiotherapy: from current concept to future perspectives *Nat. Rev. Clin. Oncol.* **9** 688–99
Keall P J *et al* 2006 The management of respiratory motion in radiation oncology report of AAPM task group 76(a) *Med. Phys.* **33** 3874–900
Kemper A R, Santago A C, Stitzel J D, Sparks J L and Duma S M 2013 Effect of strain rate on the material properties of human liver parenchyma in unconfined compression *J. Biomech. Eng.* **135** 104503
Klein S, Staring M, Murphy K, Viergever M and Pluim J 2010 elastix: a toolbox for intensity-based medical image registration *IEEE Trans. Med. Imaging* **29** 196–205
Kontaxis C, Bol G H, Lagendijk J J W and Raaymakers B W 2015 Towards adaptive IMRT sequencing for the MR-linac *Phys. Med. Biol.* **60** 2493–509
Kontaxis C, Bol G H, Stemkens B, Glitzner M, Prins F M, Kerkmeijer L G W, Lagendijk J J W and Raaymakers B W 2017 Towards fast online intrafraction replanning for free-breathing stereotactic body radiation therapy with the MR-linac *Phys. Med. Biol.* **62** 7233–48
Kopperdahl D L and Keaveny T M 1998 Yield strain behavior of trabecular bone *J. Biomech.* **31** 601–8
Korreman S S 2012 Motion in radiotherapy: photon therapy *Phys. Med. Biol.* **57** R161–R191
Lai Y S, Chen W C, Huang C H, Cheng C K, Chan K K and Chang T K 2015 The effect of graft strength on knee laxity and graft in-situ forces after posterior cruciate ligament reconstruction *PLoS One* **10** e0127293
Landau L D, Pitaevskii L P, Kosevich A M and Lifshitz E M 2012 *Theory of Elasticity: Volume 7 (Theoretical Physics)* (Oxford: Butterworth-Heinemann)
Liu Z, Deng X and zhi Wang G 2012 Accuracy validation for medical image registration algorithms: a review *Chin. Med. Sci. J.* **27** 176–81
Lu Q *et al* 2015 Hepatocellular carcinoma: stiffness value and ratio to discriminate malignant from benign focal liver lesions *Radiology* **275** 880–8
Maas S A, Ellis B J, Ateshian G A and Weiss J A 2012 FEBio: finite elements for biomechanics *J. Biomech. Eng.* **134** 011005
Maurer C, Fitzpatrick J, Wang M, Galloway R, Maciunas R and Allen G 1997 Registration of head volume images using implantable fiducial markers *IEEE Trans. Med. Imaging* **16** 447–62
Nasseri S, Bilston L E and Phan-Thien N 2002 Viscoelastic properties of pig kidney in shear, experimental results and modelling *Rheol. Acta* **41** 180–92
Oh S and Kim S 2017 Deformable image registration in radiation therapy *Radiat. Oncol. J.* **35** 101–11

Pluim J P, Muenzing S E, Eppenhof K A and Murphy K 2016 The truth is hard to make: validation of medical image registration *23rd Int. Conf. on Pattern Recognition* (IEEE) (https://doi.org/10.1109/ICPR.2016.7899978)
Raaymakers B W *et al* 2017 First patients treated with a 1.5 t MRI-linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment *Phys. Med. Biol.* **62** L41–50
Radulescu D, Peride I, Petcu L C, Niculae A and Checherita I A 2018 Supersonic shear wave ultrasonography for assessing tissue stiffness in native kidney *Ultrasound Med. Biol.* **44** 2556–68
Roach M *et al* 2011 Diagnostic and therapeutic imaging for cancer: therapeutic considerations and future directions *J. Surgical Oncol.* **103** 587–601
Rohlfing T, Maurer C, Bluemke D and Jacobs M 2003 Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint *IEEE Trans. Med. Imaging* **22** 730–41
Schreibmann E, Pantalone P, Waller A and Fox T 2012 A measure to evaluate deformable registration fields in clinical settings *J. Appl. Clin. Med. Phys.* **13** 126–39
Shamonin D 2013 Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease *Frontiers Neuroinformatics* **7**
Sirli R, Bota S, Sporea I, Jurchis A, Popescu A, Gradinaru-Tascău O and Szilaski M 2013 Liver stiffness measurements by means of supersonic shear imaging in patients without known liver pathology *Ultrasound Med. Biol.* **39** 1362–7
Sotiras A, Davatzikos C and Paragios N 2013 Deformable medical image registration: a survey *IEEE Trans. Med. Imaging* **32** 1153–90
Studholme C, Hill D and Hawkes D 1999 An overlap invariant entropy measure of 3D medical image alignment *Pattern Recognit.* **32** 71–86
Unser M 1999 Splines: a perfect fit for signal and image processing *IEEE Signal Process. Mag.* **16** 22–38
Xie L T, Yan C H, Zhao Q Y, He M N and Jiang T A 2018 Quantitative and noninvasive assessment of chronic liver diseases using two-dimensional shear wave elastography *World J. Gastroenterol.* **24** 957–70
Zachiu C, de Senneville B D, Moonen C T W, Raaymakers B W and Ries M 2018 Anatomically plausible models and quality assurance criteria for online mono- and multi-modal medical image registration *Phys. Med. Biol.* **63** 155016