Radiation Oncology

CrossMark

# Optimal number of atlases and label fusion for automatic multi-atlas-based brachial plexus contouring in radiotherapy treatment planning

Joris Van de Velde[1,2*], Johan Wouters[1], Tom Vercauteren[2], Werner De Gersem[2], Eric Achten[3],
Wilfried De Neve[2] and Tom Van Hoof[1]

## Abstract

**Background:** The present study aimed to define the optimal number of atlases for automatic multi-atlas-based brachial plexus (BP) segmentation and to compare Simultaneous Truth and Performance Level Estimation (STAPLE) label fusion with Patch label fusion using the ADMIRE® software. The accuracy of the autosegmentations was measured by comparing all of the generated autosegmentations with the anatomically validated gold standard segmentations that were developed using cadavers.

**Materials and methods:** Twelve cadaver computed tomography (CT) atlases were used for automatic multi-atlas-based segmentation. To determine the optimal number of atlases, one atlas was selected as a patient and the 11 remaining atlases were registered onto this patient using a deformable image registration algorithm. Next, label fusion was performed by using every possible combination of 2 to 11 atlases, once using STAPLE and once using Patch. This procedure was repeated for every atlas as a patient.
The similarity of the generated automatic BP segmentations and the gold standard segmentation was measured by calculating the average Dice similarity (DSC), Jaccard (JI) and True positive rate (TPR) for each number of atlases. These similarity indices were compared for the different number of atlases using an equivalence trial and for the two label fusion groups using an independent sample-*t* test.

**Results:** DSC's and JI's were highest when using nine atlases with both STAPLE (average DSC = 0,532; JI = 0,369) and Patch (average DSC = 0,530; JI = 0,370). When comparing both label fusion algorithms using 9 atlases for both, DSC and JI values were not significantly different. However, significantly higher TPR values were achieved in favour of STAPLE (*p* < 0,001). When fewer than four atlases were used, STAPLE produced significantly lower DSC, JI and TPR values than did Patch (*p* = 0,0048).

**Conclusions:** Using 9 atlases with STAPLE label fusion resulted in the most accurate BP autosegmentations (average DSC = 0,532; JI = 0,369 and TPR = 0,760). Only when using fewer than four atlases did the Patch label fusion results in a significantly more accurate autosegmentation than STAPLE.

**Keywords:** Automatic, Multi-atlas-based, Brachial plexus, Segmentation, Label fusion, Cadavers

* Correspondence: joris.vandevelde@ugent.be
Johan Wouters shared first author.
[1]Department of Anatomy, Ghent University, De Pintelaan 185, 9000 Ghent,
Belgium
[2]Department of Radiotherapy, Ghent University, De Pintelaan 185, 9000
Ghent, Belgium
Full list of author information is available at the end of the article

Van de Velde *et al. Radiation Oncology* (2016) 11:1

Page 2 of 9

## Background

Manual brachial plexus (BP) segmentation on planning computed tomography (CT) for radiation therapy treatment planning is a time-consuming and unreliable process [1]. An effective automatic BP contouring method could relieve clinicians of this tedious task and would result in a higher inter- and intra-observer reliability and accuracy of the contouring process [2]. This issue is of growing importance following the introduction of both function-sparing and adaptive intensity modulated radiation therapy (IMRT), where the number and frequency of delineation of organs at risk (OAR's) are increased [3, 4].

The BP is one of the OAR's in IMRT for head-and-neck, lung and breast cancer patients. However, in clinical practice, the BP is often not delineated during treatment planning, and when the delineation is included, the accuracy tends to be low [1]. Incorrect or absent delineations of OAR's in IMRT treatment planning however, have proven to be a main source of uncertainty in historical dose-volume effect data, which leads to the reduced performance of predictive models [5]. Moreover, when the BP radiation dose is not controlled during IMRT treatment, the possibility exists that the dose to this organ exceeds the BP tolerance dose of 66 Gy [6–8], which can potentially cause radiation-induced brachial plexopathy (RIBP). RIBP was thought to be uncommon for head-and-neck cancer patients, but recent clinical investigations have suggested that it remains underreported [9, 10].

To develop accurate automatic multi-atlas-based BP segmentations, multiple parameters must be controlled.

The first parameter is the optimal image registration and label fusion algorithm that are used. In multi-atlas-based autosegmentation strategies, several available presegmented images –called atlases– are first registered separately to the patient using deformable image registration. During the deformable image registration process a deformation vector field (DVF), describing the non-linear transformation from a presegmented image dataset to a patient image dataset, is created. Based on the computed DVF a set of delineations on the presegmented image data set are deformed on the patient image data set. The series of deformed delineations on the patient image data set are combined by the label fusion algorithm to obtain a unique and final consensus segmentation. Multiple image registration and label fusion algorithms for various organs have been compared in the literature [11–14]. However, only one publication [15] investigated BP autosegmentation. In that study, the authors concluded that multi-atlas autosegmentation can be effectively used to delineate BP on CT. However, these conclusions may be unreliable because the autosegmentation itself and also the subsequent validation procedure were based on BP gold standards that were not validated [1]. Moreover, Yang et al. [15] used the Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm to generate automatic BP contours. A described weakness of the STAPLE label fusion method is that it ignores the image data and uses only the segmentations when computing the label fusion [16]. In an attempt to counter this weakness, an additional intensity weighted label fusion method called 'Patch' fusion [17] was recently implemented in the ADMIRE® software. The ADMIRE® white paper reported that this Patch fusion outperforms the STAPLE algorithm for some anatomical structures with a very stable anatomical topography [18]. No publication has been found, however, in which the use of this label fusion algorithm was investigated for automatic multi-atlas-based BP segmentation.

A second parameter that must be controlled to obtain the maximum accuracy is the number of atlases that has to be used for multi-atlas-based BP autosegmentation. In different publications, multi-atlas-based automatic segmentation methods have proven to be more effective than single-atlas-based methods [19, 20] but the specific number of atlases to use was investigated in only a few publications [14, 21]. None of these studies provided a specific number for optimal automatic BP contouring.

The purpose of this study was to define the optimal number of atlases to use for automatic multi-atlas-based BP contouring and to compare the STAPLE algorithm with Patch label fusion using the ADMIRE® software. This was measured by comparing all of the generated automatic BP segmentations with high-quality, anatomically validated, gold standard atlases that were developed using cadavers.

## Materials and methods

To develop gold standard atlases for BP contouring, 12 cadavers (age and gender randomized) were used. The cadavers were embalmed according to Thiel because of their optimal image quality and movement capacities [22, 23]. The latter allowed for the required standardization of the scan position. Magnetic resonance imaging (MRI) of the head-and-neck region was performed to generate high-quality BP delineations that were anatomically validated by dissection. These anatomically validated, MRI-based, BP delineations were then rigidly fused to the corresponding CT to obtain BP gold standard delineations that were applicable to the radiation therapy planning system. A detailed description was provided by Van de Velde et al. [24]. This study was approved by the ethics committee of University Hospital Ghent (reference number: B67020142069), and was in compliance with the Helsinki Declaration.

Van de Velde *et al. Radiation Oncology* (2016) 11:1

Page 3 of 9

**Table 1** Average Dice similarity coefficient, Jaccard index and True positive rate per number of atlases

| Number of atlases | Samples | STAPLE | | | Patch | | |
|---|---|---|---|---|---|---|---|
| | | DSC (SD) | JI (SD) | TPR (SD) | DSC (SD) | JI (SD) | TPR (SD) |
| 2 | 660 | 0,247 (0,179) | 0,154 (0,131) | 0,188 (0,158) | 0,400 (0,157) | 0,262 (0,124) | 0,416 (0,178) |
| 3 | 660 | 0,397 (0,184) | 0,265 (0,151) | 0,373 (0,187) | 0,454 (0,157) | 0,307 (0,136) | 0,439 (0,163) |
| 4 | 3960 | 0,472 (0,171) | 0,325 (0,147) | 0,473 (0,184) | 0,477 (0,161) | 0,328 (0,141) | 0,445 (0,165) |
| 5 | 5544 | 0,482 (0,153) | 0,331 (0,132) | 0,534 (0,166) | 0,465 (0,149) | 0,316 (0,128) | 0,435 (0,150) |
| 6 | 5544 | 0,519 (0,138) | 0,362 (0,128) | 0,616 (0,155) | 0,501 (0,146) | 0,347 (0,133) | 0,465 (0,150) |
| 7 | 3960 | 0,514 (0,129) | 0,356 (0,117) | 0,658 (0,147) | 0,492 (0,144) | 0,339 (0,131) | 0,446 (0,142) |
| 8 | 1980 | 0,501 (0,120) | 0,343 (0,106) | 0,686 (0,143) | 0,501 (0,140) | 0,346 (0,127) | 0,466 (0,140) |
| 9 | 660 | 0,532 (0,102)[a] | 0,369 (0,940)[a] | 0,726 (0,127) | 0,530 (0,117)[a] | 0,370 (0,112)[a] | 0,466 (0,125) |
| 10 | 132 | 0,510 (0,100) | 0,349 (0,900) | 0,742 (0,127) | 0,524 (0,124) | 0,365 (0,116) | 0,468 (0,121) |
| 11 | 12 | 0,506 (0,940) | 0,344 (0,840) | 0,760 (0,126)[a] | 0,530 (0,122) | 0,370 (0,115) | 0,471 (0,115)[a] |

Abbreviations: *DSC* dice similarity coefficient; *JI* Jaccard index; *TPR* true positive rate
[a]Highest index values

For image registration and label fusion, the ADMIRE® software 1.10.02 (Elekta AB, Stockholm, Sweden) was used. ADMIRE® performs the segmentation of a novel subject image (here called 'patient') by using multiple pre-segmented images, which are also known as 'atlases'. The 'General' algorithm in ADMIRE® is used for the initial deformable image registration [18]. This image registration framework consists of three major steps: a linear registration and two non-linear registration steps. With each step the number of degrees of freedom increases, and is used to provide initialization for the next step.

For label fusion, 2 different algorithms in ADMIRE® are compared: the STAPLE label fusion [16] and Patch label fusion [17]. The STAPLE algorithm works with a statistical framework that simultaneously estimates the underlying 'truth' segmentation and the accuracy of each individual atlas [18]. It ignores the image data and uses only the segmentations when computing the label fusion. In contrast, the Patch algorithm considers the accuracy of the initial image registration by comparing the intensity similarity between the atlas and the patient after being aligned, to get better label fusion results. This process, is called 'intensity weighting'.

### Procedure
The present study aimed to determine the optimal number of atlases and to compare the STAPLE with the Patch label fusion algorithm for multi-atlas-based BP contouring in ADMIRE® software.

For this purpose, a leave-one-out strategy was followed. One of the 12 available cadaver CT-datasets was selected as a patient and the remaining CT-datasets, which contained the anatomically validated BP segmentation, served as atlases. All of the atlases were first registered separately onto the patient using the 'General' registration algorithm in ADMIRE®. Next, the label fusion was performed, with both STAPLE and Patch, first using every possible combination of 2 atlases. Subsequently, label fusion was repeated with a gradually increasing number of atlases, until every possible combination of 11 atlases was reached. This process was reiterated for every atlas as a patient. It resulted in 24432 combinations over the different number of atlases. A Power analysis was executed (power π = 80) to calculate the minimum sample size required for a 90 % confidence interval.

Next, for every generated 'label fused' autosegmentation, 3 similarity indices with the gold standard contour were calculated to quantify the accuracy (Table 1):

First, Dice similarity coefficient (DSC) was calculated between these 2 segmentations. The DSC measures the spatial overlap between the gold standard A and the registered image B, and is defined as DSC(A,B) = 2(A∩B)/(A + B) where ∩ is the intersection volume. The DSC is situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

We also calculated the Jaccard index (JI) as the ratio of the intersection volume and the entire union volume of the delineations: JI(A,B) = (A∩B)/(A∪B). The JI is also situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

At last, True positive rate (TPR) was measured between the gold standard BP (A) and the registered BP (B). TPR is the intersection volume of these, divided by the gold standard BP: TPR = (A∩B/A). TPR is situated between 0 and 1 with 0 indicating no inclusion and 1 indicating the total inclusion of A by B.

Finally, for each number of atlases, average DSC, JI and TPR were calculated over the different combinations.

To determine the clinically relevant optimal number of atlases, an equivalence trial was conducted [25, 26]. An equivalence trial is used to demonstrate similarity between compared groups. It uses a confidence interval

Van de Velde *et al. Radiation Oncology* (2016) 11:1

Page 4 of 9

in which equivalence is claimed when the confidence interval of the difference in outcome between compared groups is within a predetermined equivalence margin. This equivalence margin represents a clinically acceptable range of differences. For this study, an equivalence margin of 10 % was predetermined.

Only DSC and JI were appropriate as a reference for the equivalence trial, because in those indices, the most accurate segmentation will be associated with the highest index values, since both indices consider a penalty for false positive delineation area. The TPR from its side was not adequate for the equivalence trial because the highest TPR value does not necessary imply the most accurate segmentation [27], since a false positive delineation area is not penalized in this index.

DSC was chosen for equivalence trial over JI because the DSC has a linear course with an increasing correctly delineated volume and JI has not. Thus, a 10 % (= equivalence margin) increase or decrement of DSC always correlates with the same amount of increase or decrement of the correctly delineated volume [27]. Using JI conversely, the amount of correctly delineated volume associated with an increase or decrease of 10 % JI value, will vary depending on the starting value of the JI, because this index has a non-linear course. For example, an increase in JI value from 0.8 to 0.9 will result in a larger increase in percentage of correctly delineated volume than an increase from 0.2 to 0.3 [27].

Starting from the number of atlases with the maximal DSC values (reference group), the number of atlases was first gradually increased by one. If, by increasing the number of atlases each time starting from the reference group, the decrease of DSC (90 % CI) felt within the equivalence margin of 10 %, the groups were considered to be equivalent. This procedure was performed for the two label fusion groups separately [26]. Only in case of equivalent DSC values combined with significantly higher TPR values, the autosegmentation result was considered to be more accurate, because in this case the equivalence of the
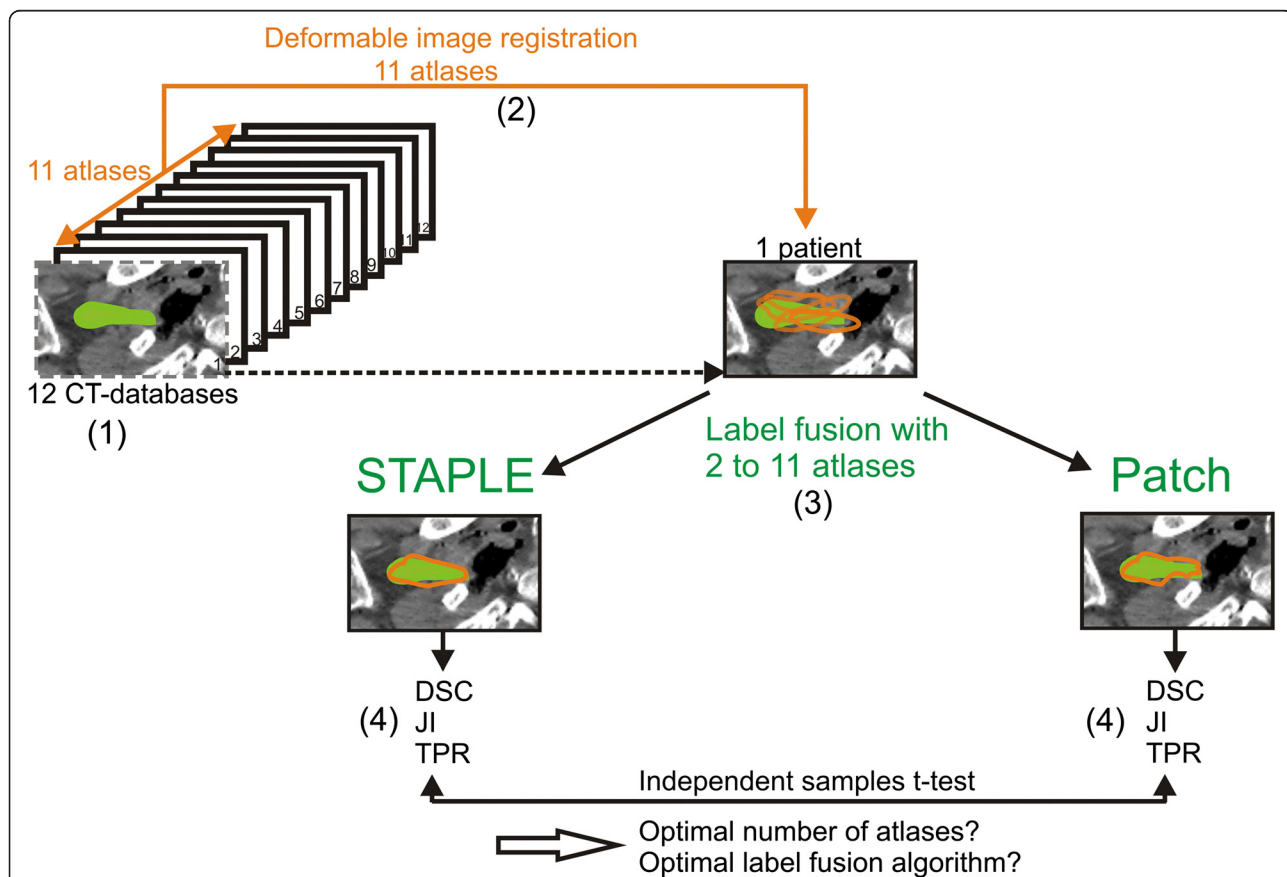


**Fig. 1** Schematic illustration of the procedure for determining the optimal number of atlases and optimal label fusion. (1) Twelve cadaver CT datasets were included, and one atlas was selected as a patient. (2) The 11 remaining atlases were used for deformable image registration on the patient. (3) Label fusion was performed with 2 up to 11 atlases, once using STAPLE and once using Patch. (4) For each number of atlases, the average Dice similarity coefficient (DSC), Jaccard index (JI) and True positive rate (TPR) were calculated for the generated contour (*orange*) with the gold standard contour (*green*). This procedure was repeated for every atlas as a patient

Van de Velde *et al. Radiation Oncology* (2016) 11:1

Page 5 of 9

DSC values indicates that the increase of the false positive delineation area, which is not penalized by TPR, was kept within bounds.

Next, the number of atlases was gradually decreased by one, starting from the reference group. If, by decreasing the number of atlases each time starting from the reference group, the decrement of the DSC values fell within the equivalence margin, the calculation time could be reduced by using a lower number of atlases without clinically relevant loss in accuracy.

Thereafter, the difference between STAPLE and Patch label fusion was determined using an independent sample *t*-test. Therefore, in the 2 label fusion groups, the similarity indices for their respective clinically relevant optimal number of atlases were compared.

## Results

The power analysis ($\pi = 80$) resulted in a sample size of 150 combinations per number of atlases needed for a 90 % confidence interval. For each number of atlases, the average DSC, JI and TPR, their standard deviations and their possible combinations (samples) are shown in Table 1 and Figs. 1, 2 and 3 for both groups.
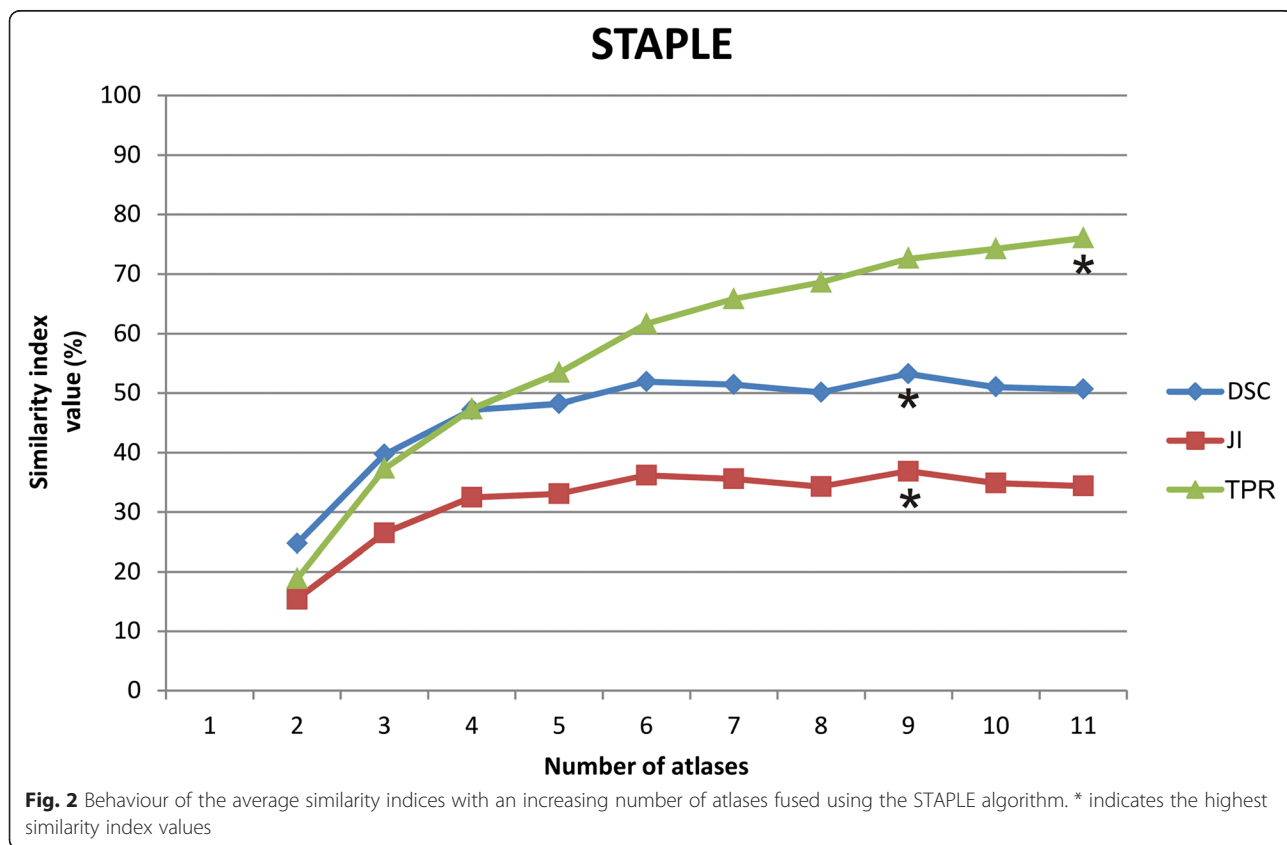
The highest average DSC and JI values were found when using 9 atlases for both STAPLE and Patch fusion (Table 1, Figs. 2 and 3). When more than 9
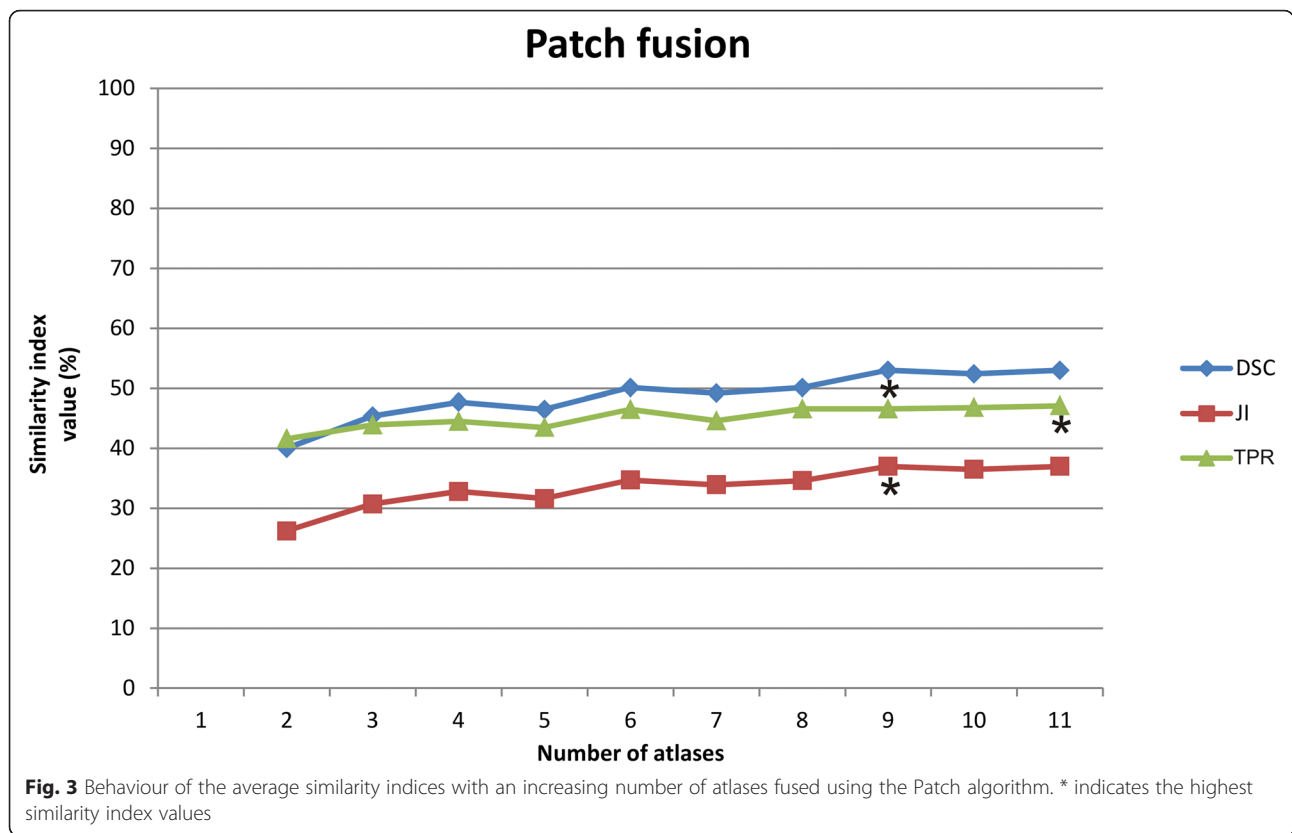
atlases were used, only TPR for STAPLE label fusion continued to increase (Fig. 2).

By increasing the number of atlases for STAPLE label fusion from 9 to 10 atlases, the decrease in DSC values still fell within the predisposed equivalence margin of 10 % (Fig. 4) but no significantly higher TPR values were achieved. When using 11 atlases, DSC were no longer equivalent to the results obtained when using 9 atlases. However, the number of possible combinations for a power of 90 (sample size of 150) was not sufficient with 10 and 11 atlases, so no definitive conclusions can be drawn concerning these number of atlases.

By decreasing the number of atlases from 9 to 6 atlases the decrease in DSC values still fell within the predisposed equivalence margin (Fig. 4). The average calculation time was reduced from 19 min to 17 min. When using fewer than 6 atlases, DSC values were no longer equivalent to the results obtained when using 9 atlases.

By increasing the number of atlases for Patch label fusion from 9 to 10 atlases, the decrease in the DSC values did fall within the predisposed equivalence margin of 10 % (Fig. 5) but no significantly higher TPR values were achieved. When the number of atlases was decreased until 8 or lower, the decrease in DSC was not within the equivalence margin (Fig. 5). Also here, the



**Fig. 2** Behaviour of the average similarity indices with an increasing number of atlases fused using the STAPLE algorithm. * indicates the highest similarity index values

Van de Velde *et al. Radiation Oncology* (2016) 11:1

Page 6 of 9



**Fig. 3** Behaviour of the average similarity indices with an increasing number of atlases fused using the Patch algorithm. * indicates the highest similarity index values

number of possible combinations for a power of 90 (sample size of 150) was not sufficient with 10 and 11 atlases.

When comparing the label fusion algorithms using the optimal number of atlases for both (9 for both STAPLE and Patch) DSC and JI values were not significantly different. However, significantly higher TPR values were achieved in favour of STAPLE ($p < 0,001$).

Only when fewer than four atlases were used, STAPLE resulted in significantly lower DSC, JI and TPR values than Patch ($p = 0,0048$).

## Discussion

The purpose of this study was to determine the optimal number of atlases for automatic multi-atlas-based BP contouring and to compare STAPLE and PATCH label fusion algorithms using the ADMIRE® software.

For STAPLE, the average DSC and JI values were maximal when using 9 atlases (Table 1). When the number of atlases was increased to 10, the DSC values remained equivalent to those obtained when using 9 atlases, but the TPR values were not significantly higher.
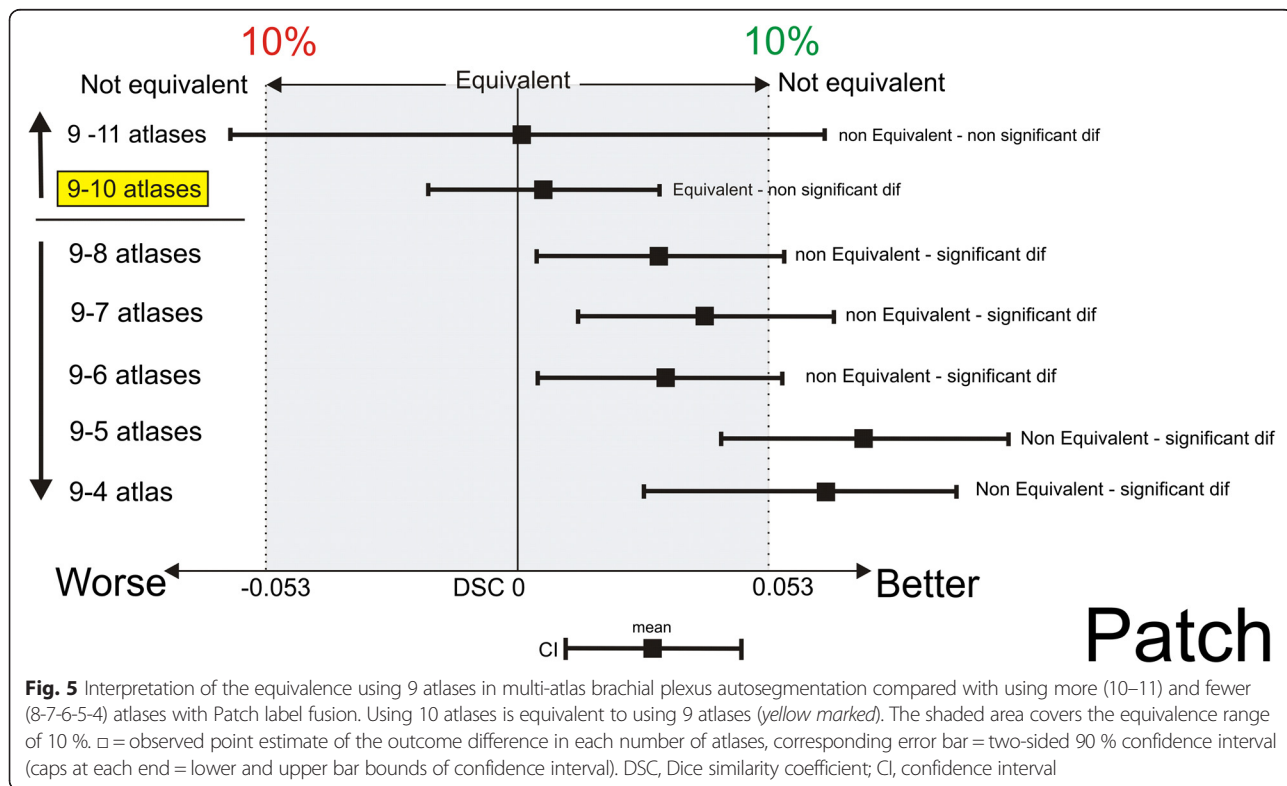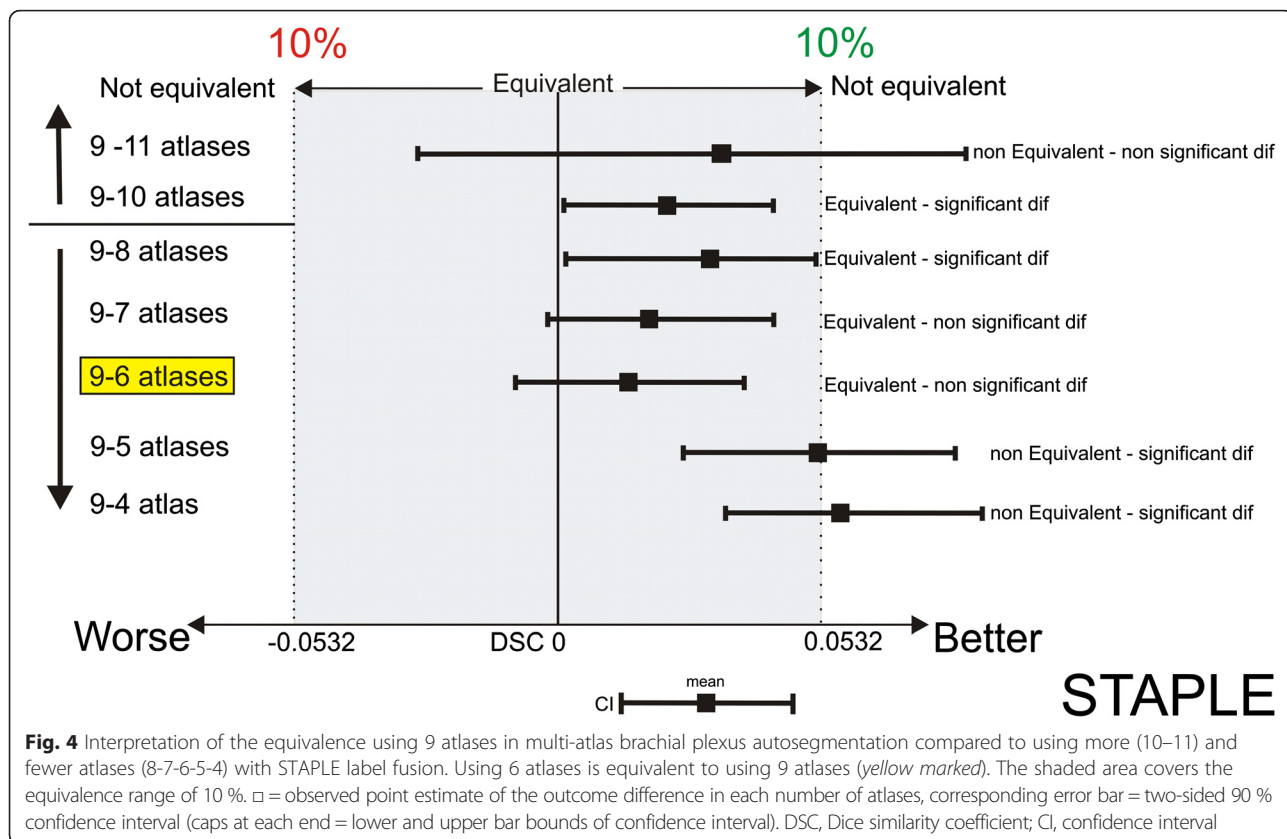
The most accurate autosegmentation results are achieved when the JI and DSC values reach their maximum value. The number of atlases where these maximum values are reached are identical for both

similarity indices. From this point on, it is only possible to achieve higher TPR values by adding more atlases. However, when the TPR values increase and the DSC and JI values decrease, the increase in the true-positive delineation area is associated with a proportionally larger increase in the false positive delineation area. This may occur because TPR only measures the increase in the true-positive delineation area and does not penalize an increase of the false positive delineation area. DSC and JI, in contrast, do penalize an increase in the false-positive delineation area.

Because the decrease in DSC from 9 to 10 atlases fell within the equivalence margin, and because the TPR values were not significantly higher, autosegmentations obtained with 10 atlases could not improve the accuracy compared to those obtained using nine atlases for STAPLE. Consequently, the optimal number of atlases is 9 for STAPLE label fusion.

In the case of limited computer calculation power, six atlases could be used for STAPLE without a clinically relevant loss of accuracy and an average time saving of 2 min.

For Patch, the DSC and JI values were also maximal at nine atlases. By increasing the number of atlases to 10, no significant increase in TPR values was achieved either, which indicated that Patch fusion with nine atlases also resulted in the most accurate autosegmentations.

Van de Velde *et al. Radiation Oncology* (2016) 11:1

Page 7 of 9



**Fig. 4** Interpretation of the equivalence using 9 atlases in multi-atlas brachial plexus autosegmentation compared to using more (10–11) and fewer atlases (8-7-6-5-4) with STAPLE label fusion. Using 6 atlases is equivalent to using 9 atlases (*yellow marked*). The shaded area covers the equivalence range of 10 %. □ = observed point estimate of the outcome difference in each number of atlases, corresponding error bar = two-sided 90 % confidence interval (caps at each end = lower and upper bar bounds of confidence interval). DSC, Dice similarity coefficient; CI, confidence interval



**Fig. 5** Interpretation of the equivalence using 9 atlases in multi-atlas brachial plexus autosegmentation compared with using more (10–11) and fewer (8-7-6-5-4) atlases with Patch label fusion. Using 10 atlases is equivalent to using 9 atlases (*yellow marked*). The shaded area covers the equivalence range of 10 %. □ = observed point estimate of the outcome difference in each number of atlases, corresponding error bar = two-sided 90 % confidence interval (caps at each end = lower and upper bar bounds of confidence interval). DSC, Dice similarity coefficient; CI, confidence interval

Van de Velde *et al. Radiation Oncology* (2016) 11:1

Page 8 of 9

To reduce the calculation time, the number of atlases cannot be decreased without a clinically relevant loss of autosegmentation accuracy (Fig. 5).

Comparing both label fusion algorithms (STAPLE and Patch) using their respective optimal number of atlases (9 for both), DSC and JI values were not significantly different. However, significantly higher TPR values were found in favour of STAPLE ($p < 0,001$). Therefore, we recommend using STAPLE label fusion with 9 atlases to obtain the most accurate autosegmentations results.

Conversely, when fewer than four atlases were used, STAPLE provided significantly less accurate results than did Patch ($p = 0.004862$). So, Patch label fusion is preferable over STAPLE when only less than 4 atlases are available.

The current study is the first to investigate the optimal number of atlases for BP autosegmentation. The optimal number of atlases for some other organs was already studied: for the nucleus caudatus Aljabar et al. [21] concluded that using eight atlases is optimal; for the hippocampus, the highest accuracy is reached with a selection of 25 atlases. Pirozzi et al. (2012) concluded that for the bladder and the femur, the optimal number of atlases was five, and that the optimal number for the prostate and rectum was four [14]. Remarkable is that in the first study [21], the number of atlases for autosegmentation of anatomically stable brain structures is higher than in the second study [14], in which anatomically variable organs were autosegmented. The opposite could be expected. The varying results of these studies only show that the optimal number of atlases is very organ-dependent and especially algorithm-dependent. So, for more general conclusions concerning the optimal number of atlases for BP autosegmentation, other algorithms also have to be investigated.

Few studies were found that compared ADMIRE® software to other autosegmentation software. Simmat et al. [11] found higher flexibility and robustness in the algorithm used in the ADMIRE® software compared with the algorithms in Iplan® [12] for the bladder, prostate and rectum. La Macchia et al. [13] found the best label fusion results using STAPLE in ADMIRE® compared with the algorithms in VelocityA® and MIM 5®, for the head-and-neck region. BP autosegmentation was not included in both studies. For general conclusions concerning the best autosegmentation software for BP autosegmentation, different autosegmentation software need to be compared in further studies.

Another limitation of the current study is that only 12 atlases were available. Hence, for the combinations with 10 and 11 atlases, the sample size was not big enough to draw definitive conclusions. To increase the statistical power and to draw definitive conclusions for the highest number of atlases, more atlases need to be included in the study. The more atlases included in the study, the more accurate the autosegmentation results will be as well, because the probability of selecting atlases that are more similar to the patient's morphotype will increase.

The dosimetric implications of optimization of label fusion and the number of atlases on radiation therapy treatment planning were not included in the investigation. Additional studies are in process to study the dosimetric impact and measure the potential benefit for patients undergoing radiation therapy treatment. Future perspectives include further increasing the accuracy of the automatic BP segmentations to a clinically acceptable level, by combining the optimal number of atlases and label fusion with an effective atlas selection strategy and including higher number of anatomically validated atlases to study the effect of using more than 11 atlases.

## Conclusion

STAPLE is preferable to Patch label fusion for multi-atlas-based BP autosegmentation. Only when fewer than four atlases are available, it's preferable to choose Patch above STAPLE.

Using nine atlases with STAPLE resulted in the most accurate BP autosegmentations. With a limited computer calculation power, the number of atlases could be decreased until 6 without a clinically relevant loss of accuracy.

### Authors' contributions
JVDV and JW contributed to the conception and design, acquisition of data and analysis, interpretation of data and drafting of the manuscript. TV was involved in the acquisition of the data, software development and drafting of the manuscript. WDS was involved in the acquisition of the data, software development and drafting of the manuscript. EA contributed to the conception and design and drafting of the manuscript. WDN contributed to the conception and design and drafting of the manuscript. TVH coordinated the investigation, was involved in the acquisition of data and analysis, interpretation of data and drafting of the manuscript. All authors read and approved the final manuscript.

### Author details
[1]Department of Anatomy, Ghent University, De Pintelaan 185, 9000 Ghent, Belgium. [2]Department of Radiotherapy, Ghent University, De Pintelaan 185, 9000 Ghent, Belgium. [3]Department of Radiology, Ghent University, De Pintelaan 185, 9000 Ghent, Belgium.

Van de Velde *et al. Radiation Oncology* (2016) 11:1

Page 9 of 9

## References

1. Van de Velde J, Vercauteren T, De Gersem W, Wouters J, Vandecasteele K, Vuye P, et al. Reliability and accuracy assessment of radiation therapy oncology group-endorsed guidelines for brachial plexus contouring. Strahlenther Onkol. 2014;190:628–35.
2. Stapleford LJ, Lawson JD, Perkins C, Edelman S, Davis L, McDonald MW, et al. Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. Int J Radiat Oncol Biol Phys. 2010;77:959–66.
3. Thomson D, Boylan C, Liptrot T, Aitkenhead A, Lee L, Yap B, et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. Radiat Onco. 2014;9(1):173.
4. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. Med Phys. 2014;41(5):050902.
5. Brouwer CL, Steenbakkers RJ, Van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D variation in delineation of head and neck organs at risk. Radiat Oncol. 2012;7:32.
6. Truong M, Romesser PB, Qureshi MM, Orlina L, Willins JD. Dose Volume Histogram Analysis of Routine Brachial Plexus Contouring for Head and Neck Intensity Modulated Radiation Therapy. Int J Radiat Oncol Biol Phys. 2011;81:S532–2.
7. Hall WH, Guiou M, Lee NY, Dublin A, Narayan S, Vijayakumar S, et al. Development and validation of a standardized method for contouring the brachial plexus: preliminary dosimetric analysis among patients treated with imrt for head-and-neck cancer. Int J Radiat Oncol Biol Phys. 2008;72:1362–7.
8. Kong FM, Ritter T, Quint DJ, Senan S, Gaspar LE, Komaki RU, et al. Consideration of dose limits for organs at risk of thoracic radiotherapy: atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. Int J Radiat Oncol Biol Phys. 2011;81:1442–57.
9. Chen AM, Hall WH, Li J, Beckett L, Farwell DG, Lau DH, et al. Brachial Plexus-Associated Neuropathy After High-Dose Radiation Therapy for Head-and-Neck Cancer. Int J Radiat Oncol Biol Phys. 2012;84:165–9.
10. Mutter RW, Lok BH, Dutta PR, et al. Constraining the brachial plexus does not compromise regional control in oropharyngeal carcinoma. Radiat Oncol. 2013;8:173.
11. Simmat I, Georg P, Georg D, Birkfellner W, Goldner G, Stock M. Assessment of accuracy and efficiency of atlas-based autosegmentation for prostate radiotherapy in a variety of clinical conditions. Strahlenther Onkol. 2012;188:807–13.
12. Noblet V, Heinrich C, Heitz F, Armspach JP. Retrospective evaluation of a topology preserving non-rigid registration method. Med Image Anal. 2006;10:366–84.
13. La Macchia M, Fellin F, Amichetti M, Cianchetti M, Gianolini S, Paola V, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. Radiat Oncol. 2012;7:160.
14. Pirozzi S, Horvat M, Piper J, et al. Atlas-based Segmentation: Evaluation of a Multi-Atlas Approach for Lung Cancer. Med Phys. 2012;39:3677.
15. Yang J, Amini A, Williamson R, et al. Automatic contouring of brachial plexus using a multi-atlas approach for lung cancer radiation therapy. PRO. 2013;3:139–47.
16. Warfield SK, Zou KH, Wells WM. Validation of image segmentation by estimating rater bias and variance. Philos Trans A Math Phys Eng Sci. 2008;366:2361–75.
17. Coupe P, Manjon JV, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. Neuroimage. 2011;54:940–54.
18. Han X, Elekta AB. A Locally Adaptive, Intensity-Based Label Fusion Method for Multi-Atlas Auto-Segmentation. MO: Maryland Heights; 2012.
19. Teguh DN, Levendag PC, Voet PWJ, Al-Mamgani A, Han X, Wolf TK, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. Int J Radiat Oncol Biol Phys. 2011;81:950–7.
20. Anders LC, Stieler F, Siebenlist K, Schafer J, Lohr F, Wenz F. Performance of an atlas-based autosegmentation software for delineation of target volumes for radiotherapy of breast and anorectal cancer. Radiother Oncol. 2012;102:68–73.
21. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage. 2009;46:726–38.
22. Thiel W. Die Konservierung ganzer leichen in natürlichen farben. Ann Anat. 1992;174:185–95.
23. De Crop A, Bacher K, Van Hoof T, Smeets PV, Smet BS, Vergauwen M, et al. Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study. Radiology. 2012;262:298–304.
24. Van de Velde J, Audenaert E, Speleers B, Vercauteren T, Mulliez T, Vandemaele P, et al. An Anatomically Validated Brachial Plexus Contouring Method for Intensity Modulated Radiation Therapy Planning. Int J Radiat Oncol Biol Phys. 2013;87:802–8.
25. Veronesi U, Orecchia R, Maisonneuve P, Viale G, Rotmensz N, Sangalli C, et al. Intraoperative radiotherapy versus external radiotherapy for early breast cancer (ELIOT): a randomised controlled equivalence trial. Lancet Oncology. 2013;14:1269–77.
26. Ahn S, Park SH, Lee KH. How to Demonstrate Similarity by Using Noninferiority and Equivalence Statistical Testing in Radiology Research. Radiology. 2013;267:328–38.
27. Van de Velde J, Wouters J, Vercauteren T, De Gersem W, Duprez F, De Neve W et al. Morphometric Atlas Selection for Automatic Brachial Plexus Segmentation. Int J Radiat Oncol Biol Phys. 2015. doi:10.1016/j.ijrobp.2015.02.045.