Radiotherapy to the head and neck is challenging due to complex anatomy and large number of organs at risk (OARs). Current radiotherapy techniques such as Intensity Modulated Radiotherapy (IMRT) increase dose conformity allowing improved loco-regional tumour control as well as reduced normal tissue effects [1,2]. To fully exploit the advantages of IMRT, accurate and consistent target delineation is required. Manual target volume and OAR delineation are affected by clinician variability [3]. Minimising interobserver variability will improve the accuracy of the dose delivered, maximise tumour control, limit toxicities and increase knowledge of organ at risk (OAR) dose [4–6]. Methods to standardise OAR volumes have been developed including superior imaging techniques, peer review and the development of contouring atlases [7].

Contouring atlases can help standardise volumes, reduce inter-observer variability and improve normal tissue sparing in daily clinical practice [8–11]. Atlases agreed by an expert panel may reduce inconsistencies between radiotherapy centres and facilitate multi-institutional clinical trials [12]. There are a number of atlases for head and neck cancer: the current Danish Head and Neck multi-institutional clinical trials[12]. To fully exploit the advantages of IMRT, accurate and consistent target delineation is required. Manual target volume and OAR delineation are affected by clinician variability [3]. Minimising interobserver variability will improve the accuracy of the dose delivered, maximise tumour control, limit toxicities and increase knowledge of organ at risk (OAR) dose [4–6]. Methods to standardise OAR volumes have been developed including superior imaging techniques, peer review and the development of contouring atlases [7].

Contouring atlases can help standardise volumes, reduce inter-observer variability and improve normal tissue sparing in daily clinical practice [8–11]. Atlases agreed by an expert panel may reduce inconsistencies between radiotherapy centres and facilitate multi-institutional clinical trials [12]. There are a number of atlases for head and neck cancer: the current Danish Head and Neck
Cancer Group (DAHANCA), European Organization for Research and Treatment of Cancer (EORTC) and Radiation Therapy Oncology Group (RTOG) [12]. One limitation of current published atlases is the absence of delineation guidelines for the masticatory muscles.

Trismus is defined as a maximum inter-incisor distance of \(< 35\text{ mm}\) and is caused by impaired function of the masticatory muscles [13]. Trismus can manifest in poor dental hygiene, impaired chewing, malnutrition and psychological difficulties including low self-esteem, depression and suicidal intentions, which all reduce health-related quality-of-life [14,15]. Clinical assessment of patients is challenging due to a restricted ability to assess disease status. There are several patient, tumour and treatment related factors for trismus of which radiotherapy is a known contributor with an incidence in advanced oropharyngeal cancers of 35–55% [16–20]. Mouth opening is a complex action controlled by the synergistic actions of the paired muscles of mastication. These include: medial and lateral pterygoids (MP, LP), masseter (M), temporalis (T) as well as the temporo-mandibular joint (TMJ) [21]. The origin, insertion and function of each of the muscles of mastication are summarised in Supplementary Tables 1a and 1b.

Identifying the masticatory apparatus as an OAR with a view to avoidance planning will aim to reduce toxicities and improve quality-of-life. Dosimetric studies showed a relationship between the severity of trismus with dose and volume of muscle treated [22]. Despite this there is no standardised defined OAR or dose threshold for the masticatory muscles for radiotherapy planning [23]. Within the current literature, inconsistences exist regarding proposed dose parameters as summarised in Supplementary Table 2. For example, the largest most recent study by Rao et al of 421 patients suggested limiting the high dose volume of the ipsilateral MP to \(V_{68}\text{ Gy} < 10\text{ cm}^3\) [16].

Few studies have evaluated the use of delineation guidelines to improve interobserver variability in contouring OARs. A paper by Brouwer et al. of 6 head and neck clinicians showed poor compliance with delineation guidelines for the spinal cord, parotid and submandibular glands was associated with an increase in interobserver variability [24]. Currently there are no standardised delineation guidelines for contouring the masticatory muscles [12]. Accurate delineation of these muscles with a validated atlas is a prerequisite for high quality radiotherapy planning to improve consistency, standardise contours and reduce radiation related trismus. This study aimed to evaluate a novel muscles of mastication atlas to aid clinician contouring, reduce interobserver variability, support training and the development of multi-institutional clinical trials.

Methods

A muscles of mastication atlas was developed by a multi-disciplinary expert team consisting of a consultant radiologist, maxillo-facial surgeon and clinical oncologists. Using the Pinnacle (Pinnacle version 9.6, Philips Radiation Oncology Systems, Andover, MA) treatment planning system, the muscles of mastication (MP, LP, M, T) and TMJ were contoured on computed tomography (CT) slices. All muscles were delineated using the soft tissue window with the exception of the TMJ which was contoured on a bone window. The contours were extracted as DICOM files and converted into an app using in-house software. The atlas app is shown in Fig. 1, with a link attached [www.bit.ly/trismusatlas](http://www.bit.ly/trismusatlas) (to access the webpage please open with google chrome version 56 or opera version 43). Included in the atlas app is a table explaining the anatomical boundaries of each component of the muscles of mastication.

Seven head and neck clinicians (five consultants and two trainees not included in the multi-disciplinary team) delineated the paired muscles of mastication on randomly selected CT scans from five patients without the atlas. After a minimum gap of five weeks each clinician was given the atlas and re-contoured the same structures on the same five CT scans. Contours were created for each patient by the same multi-disciplinary team that produced the atlas and used as the reference. In-house software was used to compare clinician-drawn structures with the reference. Dice similarity coefficient (DSC), mean distance to agreement (DTA) and the centre of mass...
difference (COM) were evaluated and compared to the reference for each volume without and with the atlas. DTA was calculated by measuring the distance from each point on the reference surface to the closest point on the clinician-drawn surface and combining into a DTA histogram. The mean DTA was then calculated from this DTA histogram. The mean and standard deviation (SD) across all patients were compared without and with the atlas to assess inter-observer variability. Comparison was also performed split by training grade. Standard deviation maps were produced to illustrate the variation in structure delineation between clinicians at each voxel without and with the atlas. The Standard deviation (SD) per voxel is calculated in 3D by combining all clinician contours. All voxels inside a contour are given a value of 1 and all outside the contour are given a value of 0. The standard deviation of each voxel is calculated to illustrate regions in which there is little agreement between clinicians: the larger the standard deviation the larger the disagreement.

### Statistical analysis

Analysis was performed using Graph pad prism version 6 (Graph pad software) and Microsoft Office Excel 2010. A paired t-test was used to compare mean DSC, mean DTA and distance to COM without and with the atlas. Statistical significance was defined as $p \leq 0.05$.

### Results

The median (range) time between contouring without and with the atlas across all seven clinicians was 66 (35–145) days. Using the atlas there was an increase in the mean delineated volumes for all muscles excluding the TMJ (Table 1). The SD of the contoured volumes significantly reduced using the atlas for the MP ($p = 0.01$), T ($p = 0.05$) and TMJ ($p < 0.01$). The difference in distance between the COM and SD reduced significantly in all directions.

### Table 1

Comparison of the contoured volumes of the muscles of mastication without and with the atlas.

<table>
<thead>
<tr>
<th>Volume</th>
<th>Mean ± SD (cm$^3$)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No atlas</td>
<td>Atlas</td>
</tr>
<tr>
<td>Lateral pterygoids</td>
<td>6.7 ± 1.5</td>
<td>6.8 ± 1.2</td>
</tr>
<tr>
<td>Medial pterygoids</td>
<td>7.7 ± 1.9</td>
<td>8.5 ± 1.3</td>
</tr>
<tr>
<td>Masseters</td>
<td>19.4 ± 2.1</td>
<td>20.1 ± 1.7</td>
</tr>
<tr>
<td>Temporalis'</td>
<td>20.3 ± 6.6</td>
<td>28.0 ± 4.3</td>
</tr>
<tr>
<td>TMJs</td>
<td>1.7 ± 1.1</td>
<td>1.3 ± 0.3</td>
</tr>
</tbody>
</table>

Abbreviations: SD = standard deviation; TMJ = temporo-mandibular joint.
with the atlas for the T: anterior-posterior 2.1 ± 1.4 vs 4.7 ± 4.7 mm, \( p = 0.03 \); left–right 4.5 ± 3.0 vs 8.7 ± 5.9 mm, \( p = 0.03 \); superior-inferior 3.4 ± 2.8 vs 7.0 ± 4.7 mm, \( p = 0.05 \). No significant difference in the COM distance was observed with the atlas for the LP, MP and M.

Mean DSC significantly improved using the atlas for the LP (0.8 ± 0.1 vs 0.8 ± 0.1 \( p < 0.01 \)), MP (0.7 ± 0.2 vs 0.7 ± 0.2, \( p < 0.01 \)), T (0.7 ± 0.2 vs 0.8 ± 0.1, \( p < 0.01 \)) and TMJ (0.6 ± 0.2 vs 0.8 ± 0.1, \( p < 0.01 \)). No significant improvement in mean DSC was observed using the atlas for the M (0.9 ± 0.1 vs 0.9 ± 0.0, \( p = 0.27 \)) (Fig. 2).

Mean DTA improved using the atlas for all muscles, reaching significance for the MP (3.5 ± 4.1 mm vs 3.0 ± 3.8, \( p = 0.01 \)), M (1.4 ± 0.4 vs 1.2 ± 0.4, \( p < 0.01 \)), T (5.4 ± 5.6 vs 1.6 ± 1.7 mm, \( p < 0.01 \)) and TMJ (1.7 ± 1.2 vs 0.9 ± 0.7 mm, \( p < 0.01 \)), see Fig. 3. Using the atlas, the mean DTA improved for the LP but the variability increased, however this was not significant (1.5 ± 0.5 vs 1.4 ± 0.7 mm, \( p = 0.09 \)).

Fig. 4 illustrates standard deviation maps on representative slices for the M and T for a single patient. Regions in which there is variation between clinician contours are shown in varying degrees of blue – the darker the shade of blue the larger the variation. The atlas reduced the variation between clinicians for the T particularly at the cranial and caudal aspects of the muscle. The reduction in variability with the atlas was smaller for the M.

Table 2 shows the analysis of the mean ± SD DTA without and with the atlas performed according to clinician training grade. An improvement in mean DTA using the atlas was observed by the trainees across all masticatory muscles, with the largest improvement and reduction in variability noted for the T (4.3 ± 7.1 vs 1.2 ± 0.4 mm, \( p = 0.06 \)) and TMJ (2.1 ± 0.7 vs 0.8 ± 0.3 mm, \( p < 0.01 \)).

**Discussion**

This prospective study is the first to test the feasibility of a novel atlas of muscles of mastication for contouring in head and neck radiotherapy. Feasibility was defined as a reduction in interobserver variability. This study showed the atlas: (i) improved spatial overlap and alignment of contours for the LP, MP, T and TMJ; and (ii) improved consistency in contouring of all masticatory muscles by the trainees.

Radiation induced trismus is a significant cause of treatment related morbidity [25]. The main application of contouring the masticatory apparatus as an avoidance structure is for tumours not infiltrating the muscles in order to spare normal healthy tissue. Clinician variation in OAR contouring may lead to over dosage of the muscles of mastication and as a consequence trismus and poor quality of life. There is overlap between the development of trismus and other health-related quality of life variables. In a paper...
by Lee et al functional deficits such as taste disturbance, pain, dry mouth and social functioning were increased in patients with trismus [26].

The masticatory muscles are not routinely contoured as avoidance structures. This is partly due to insufficient understanding and consensus on dose–response relationships and a lack of standardized volumes for radiotherapy planning [13,16,18,27]. Correct identification of the TMJ as an avoidance structure is important to prevent and delay development of trismus, particularly in nasopharynx cancers [28,29]. Improving the standardisation of contouring the MP, T and TMJ will enhance consistency in the study of dose–response relationships [21].

In our study use of the atlas was shown to significantly reduce the mean and SD for DSC and DTA suggesting an improvement in interobserver variability for the MP, T and TMJ. Although the improvement in mean DTA was clinically significant, there was only a small improvement in SD (1.2 ± 0.4 vs 1.4 ± 0.4 mm) with the atlas observed for the M. Clinicians are more experienced in contouring the M and anatomically it is easier to define. Using the atlas, consistency in contouring did not significantly improve for the LP. An increase in the mean volume was observed for all contoured muscles using the atlas excluding the TMJ. The TMJ is defined by bone limits and was thus contoured using the atlas on the CT bone window. The reduction in TMJ volume and significant improvement in contouring consistency (DTA and DSC) may be explained by the superior visualisation of the bone and joint structures. Using the atlas, improvements in mean DSC and mean DTA were noted for the TMJ (0.6 ± 0.2 vs 0.8 ± 0.1, p < 0.01) and (1.7 ± 1.2 vs 0.9 ± 0.7 mm, p < 0.01) respectively. The improvement in mean DTA is a large significant change for a small, flat structure. Furthermore, the increase in DSC was despite a reduction in volume with the atlas. DSC is correlated with volume so the large increase observed here was a direct result of the atlas. There is no consensus on which metrics should be used for assessing inter-observer variation, and so for the present study both a volume-based (DSC) and surface-based (DTA) approach were used.

The atlas significantly improved consistency of contours at the muscle extremities in particular the T as shown in Fig. 4. There is currently no established dose–response relationship for the T muscle and no consensus as to which masticatory muscles are important in the development of trismus. The T muscle was identified by the MDT team as being an important masticatory muscle and therefore included in the atlas.

Using the atlas, there was a reduction in variability of contours of all muscles by the trainees, in contrast with the consultants that only more consistently contoured the T and TMJ. The improvement in DTA may be explained by the superior visualisation of the bone and joint structures. Using the atlas, improvements in mean DSC and mean DTA were noted for the TMJ (0.6 ± 0.2 vs 0.8 ± 0.1, p < 0.01) and (1.7 ± 1.2 vs 0.9 ± 0.7 mm, p < 0.01) respectively. The improvement in mean DTA is a large significant change for a small, flat structure. Furthermore, the increase in DSC was despite a reduction in volume with the atlas. DSC is correlated with volume so the large increase observed here was a direct result of the atlas. There is no consensus on which metrics should be used for assessing inter-observer variation, and so for the present study both a volume-based (DSC) and surface-based (DTA) approach were used.

The atlas significantly improved consistency of contours at the muscle extremities in particular the T as shown in Fig. 4. There is currently no established dose–response relationship for the T muscle and no consensus as to which masticatory muscles are important in the development of trismus. The T muscle was identified by the MDT team as being an important masticatory muscle and therefore included in the atlas.

Using the atlas, there was a reduction in variability of contours of all muscles by the trainees, in contrast with the consultants that only more consistently contoured the T and TMJ. The improvement

---

Table 2
Comparison of DTA without and with the atlas between trainee and consultant grade.

<table>
<thead>
<tr>
<th></th>
<th>Trainees</th>
<th></th>
<th>P value</th>
<th>Consultants</th>
<th></th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD DTA (mm)</td>
<td>Atlas</td>
<td>P value</td>
<td>Mean ± SD DTA (mm)</td>
<td>Atlas</td>
<td>P value</td>
</tr>
<tr>
<td>Lateral pterygoids</td>
<td>1.1 (0.5)</td>
<td>1.1 (0.40)</td>
<td>0.06</td>
<td>1.5 (0.6)</td>
<td>1.5 (0.7)</td>
<td>0.68</td>
</tr>
<tr>
<td>Medial pterygoids</td>
<td>3.4 (3.9)</td>
<td>2.3 (3.1)</td>
<td>0.34</td>
<td>3.5 (4.2)</td>
<td>3.5 (4.2)</td>
<td>0.97</td>
</tr>
<tr>
<td>Masseters</td>
<td>1.3 (0.4)</td>
<td>1.2 (0.3)</td>
<td>0.46</td>
<td>1.4 (0.4)</td>
<td>1.2 (0.4)</td>
<td>0.12</td>
</tr>
<tr>
<td>Temporals</td>
<td>4.3 (7.1)</td>
<td>1.2 (0.4)</td>
<td>0.06</td>
<td>5.9 (4.8)</td>
<td>1.8 (2.0)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>TMJs</td>
<td>2.1 (0.7)</td>
<td>0.8 (0.3)</td>
<td>&lt;0.01</td>
<td>1.6 (1.4)</td>
<td>1.0 (0.8)</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Abbreviations: DTA = distance to agreement; SD = standard deviation; TMJ = temporo-mandibular joint.

---

Please cite this article in press as: Hague C et al. Use of a novel atlas for muscles of mastication to reduce inter observer variability in head and neck radiotherapy contouring. Radiother Onclol (2018), https://doi.org/10.1016/j.radonc.2018.10.030
in consistency across all muscles by the trainees implies the benefit of the atlas as an educational tool for trainees including dosimetrists. The atlas may also be used by other radiotherapy centres to improve consistency, knowledge and establish collaborations to aid the development of multi-institutional clinical trials. Development of NTCP models can be facilitated by generating agreement in dose constraint parameters, facilitated by a greater consistency in contouring the muscles of mastication. The reduction in variability in contouring the muscles of mastication may translate into a reduction in variability in reported dose to these structures [30]. It is beyond the scope of the present study however to determine the dosimetric effects of reduced clinician inter-observer variation. Future studies to integrate the atlas into an auto-contouring model to reduce inter and intraobserver variability and minimise time constraints should also be considered [31]. Improving consistency of contours of the MP, T and TMJ will help standardise volumes, develop more precise dosimetric parameters which can be implemented into avoidance radiotherapy planning to potentially improve radiation related trismus and quality of life [21].

Whilst this is the first paper to our knowledge that uses a novel atlas for the muscles of mastication to evaluate interobserver variability, the study did not explore intraobserver variability or time constraints. The study only involved CT scan images of five patients, however as seven clinicians contoured each plan, a good measure of interobserver variability was obtained. Established dose–response relationships will be facilitated by a more consistent clinician approach to contouring the masticatory muscles. Future studies with greater consistency in contouring and larger numbers are required to further evaluate dose constraints.

A novel atlas has been developed to contour the masticatory muscles during head and neck radiotherapy planning. The atlas has been shown to significantly reduce interobserver variability for the MP, T and TMJ. The atlas could be considered as an education tool to improve knowledge amongst trainees and provide contouring consistency to aid the development of multi-institutional clinical trials. The atlas has been developed into an app for wider distribution amongst radiotherapy centres. Reducing interobserver variability and standardising treatment volumes will improve the accuracy of avoidance planning and potentially reduce radiation related trismus.

Conflicts of interest
None.

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2018.10.030.

References