ACTA ONCOLOGICA
2022, VOL. 61, NO. 6, 672–679
https://doi.org/10.1080/0284186X.2022.2036367

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Check for updates

International assessment of interobserver reproducibility of flap delineation in head and neck carcinoma

Arnaud Beddok[a,b], Leslie Guzene[c], Alexandre Coutte[c], David Thomson[d], Sue S. Yom[e], Valentin Calugaru[a], Eivind Blais[f], Olivier Gilliot[f], Séverine Racadot[g], Yoann Pointreau[h], June Corry[i], Kenneth Jensen[j], Sandro Porceddu[k], Nazim Khalladi[l], Vianney Bastit[m], Audrey Lasne-Cardon[m], Pierre-Yves Marcy[n], Florent Carsuzaa[o], Christophe Nioche[b], Jean Bourhis[p], Julia Salleron[q] and Juliette Thariat[l,r]; on behalf of the GORTEC

[a]Department of Radiation Oncology, Institut Curie, Paris – Orsay, France; [b]Laboratoire d'Imagerie Translationnelle en Oncologie (LITO), U1288 Université Paris Saclay/Inserm/Institut Curie, Orsay, France; [c]Department of Radiation Oncology, University Hospital of Amiens, Amiens, France; [d]Department of Radiation Oncology, The Christie NHS Foundation Trust, Manchester, United Kingdom; [e]Department of Radiation Oncology, University of California San Francisco, USA; [f]Department of Radiation Oncology, Polyclinique Marzet, Pau, France; [g]Department of Radiation Oncology, Centre Léon Bérard Lyon, France; [h]Department of Radiation Oncology, Centre Jean Bernard, Le Mans, France; [i]Department of Radiation Oncology, GenesisCare. St Vincent's Hospital, Fitzroy, Australia; [j]Department of Radiation Oncology, Aarhus University Hospital, Aarhus, Danemark; [k]Department of Radiation Oncology, Princess Alexandra Hospital Southside Clinical Unit, Australia; [l]Department of Radiation Oncology, Centre François Baclesse, Caen, France; [m]Department of Head and Neck Surgery, Centre François Baclesse, Caen, France; [n]Department or Radiology, Polyclinique les Fleurs, Ollioules, France; [o]Department of Head and Neck Surgery, University Hospital of Poitiers, Poitiers, France; [p]Department of Radiation Oncology, University Hospital of Vaudois, Lausanne, Swiss; [q]Department of Statistics, Lorraine Cancer Institute, Vandoeuvre-lès-Nancy, France; [r]Laboratoire de physique Corpusculaire IN2P3/ENSICAEN/ CNRS UMR 6534 - Normandie Université, Caen, France

ABSTRACT

Background: Several reports have suggested that radiotherapy after reconstructive surgery for head and neck cancer (HNC), could have deleterious effects on the flaps with respect to functional outcomes. To predict and prevent toxicities, flap delineation should be accurate and reproducible. The objective of the present study was to evaluate the interobserver variability of frequent types of flaps used in HNC, based on the recent GORTEC atlas.

Materials and methods: Each member of an international working group (WG) consisting of 14 experts delineated the flaps on a CT set from six patients. Each patient had one of the five most commonly used flaps in HNC: a regional pedicled pectoralis major myocutaneous flap, a local pedicled rotational soft tissue facial artery musculo-mucosal (FAMM) (2 patients), a fasciocutaneous radial forearm free flap, a soft tissue anterolateral thigh (ALT) free flap, or a fibular free flap. The WG's contours were compared to a reference contour, validated by a surgeon and a radiologist specializing in HNC. Contours were considered as reproducible if the median Dice Similarity Coefficient (DSC) was > 0.7.

Results: The median volumes of the six flaps delineated by the WG were close to the reference contour value, with approximately 50 cc for the pectoral, fibula, and ALT flaps, 20 cc for the radial forearm, and up to 10 cc for the FAMM. The volumetric ratio was thus close to the optimal value of 100% for all flaps. The median DSC obtained by the WG compared to the reference for the pectoralis flap, the FAMM, the radial forearm flap, ALT flap, and the fibular flap were 0.82, 0.40, 0.76, 0.81, and 0.76, respectively.

Conclusions: This study showed that the delineation of four main flaps used for HNC was reproducible. The delineation of the FAMM, however, requires close cooperation between radiologist, surgeon and radiation oncologist because of the poor visibility of this flap on CT and its small size.

ARTICLE HISTORY
Received 11 November 2021
Accepted 23 January 2022

KEYWORDS
Interobserver variability; head and neck carcinoma; radiation therapy

Introduction

The standard use of postoperative radiotherapy (PORT) for resected stage III/IV head and neck carcinoma (HNC) is based on retrospective and prospective studies. These have demonstrated better locoregional control after PORT than after surgery alone for patients with pathologic risk factors for recurrence, such as close/positive margins, tumor size, number of involved nodes, extranodal extension, anatomic location, perineural invasion, compared with patients without risk factors [1–3]. In contrast to definitive (tumor-in-place) irradiation, for which there are standard recommendations to guide target volume delineation based on clinical, imaging, and pathology reporting [4], there are few recommendations for target volume delineation for PORT [5]. The postoperative clinical target volume (CTV) is classically defined to include

CONTACT Juliette Thariat jthariat@gmail.com Department of radiotherapy, Centre François Baclesse, Caen, France; Laboratoire de physique Corpusculaire IN2P3/ENSICAEN/CNRS UMR 6534 - Normandie Université, Caen, France

Supplemental data for this article can be accessed here.

tissues that are at risk of tumor recurrence, i.e. tissues in contact with the resected tumor (classically called the tumor bed) and potential microscopic tumor extensions. For large HNC, reconstructive surgery is frequently performed using an autologous flap, harvested from the patient donor site and transferred to the tumor bed to compensate for the loss of substance, both for correct functional results (phonation, mastication, swallowing), and for cosmetic purposes [6,7]. The delineation of the postoperative CTV is, in this case, even more complicated, because it is not known if the flap should be included in its entirety in the CTV, or if on the contrary the major part of the flap should be protected [8]. Indeed, several surgical reports have suggested that PORT could have deleterious effects on the flaps with respect to functional outcomes [9,10]. It is not mentioned if in these studies the flap was delineated, and if a dose constraint was applied on this organ at risk (OAR). There are therefore many uncertainties, and need of studies evaluating dose- effects regarding functional outcome. In a recent study, members of the French 'Groupe des Oncologues Radiothérapeutes Tête et Cou' (GORTEC) proposed, for the first time, an atlas for the delineation of the most frequently used flaps in reconstructive surgery for head and neck cancer [11]. Based on this atlas, the objective of the present study was to evaluate the interobserver variability (IOV) of flap delineation for different types of flaps used in reconstructive surgery for HNC. This study is part of the Xflap program and is being used in the multicentric 'Xflap' study (registered as a non-interventional study under GDPR-compliant CNIL/MR004 2214228 v0 number on July 3rd, 2019).

Material and methods

Creation of the working group

An international working group was created to evaluate and improve flap delineation. The working group consisted of 14 radiation oncologists, one radiologist, and three surgeons. All of them were specialized in HNC diagnosis and treatment. General principles guiding the activities of the working group were: (1) to integrate as accurately as possible anatomic knowledge and previously defined surgical and radiological guidelines from the before-mentioned atlas [11], and (2) to minimize variations in interpretation of the guidelines between radiation oncologists by refining the description of the different flaps. The delineation process was based on the operative report, to identify the type of flap and its components (fat, fascia, muscle, skin, bone).

Construction of the data sets

During step 1, each member of the working group delineated the flaps on a CT set from five patients. Each patient had been surgically treated for HNC and had one of the five most commonly used flaps in reconstructive surgery for these cancers (Table 1): a regional pedicled pectoralis major myocutaneous flap (patient A), a local pedicled rotational soft tissue facial artery musculo-mucosal (FAMM, patient B), a fasciocutaneous radial forearm free flap (patient C), a soft tissue anterolateral thigh (ALT) free flap (patient D), and a fibular free flap (patient E) (Table 1). This study is a part of the X flap program which has received IRB approval (GDPR MR004 2214228v0). Patients who had not indicated their non-opposition to the use of data concerning them were excluded. For each patient, imaging, surgical, and pathology reports were provided in English to the working group observers to assist them in delineating the flaps. CT images (2 mm slice thickness) were performed with the patient in supine position on a planning multidetector spiral scanner and a thermoformed mask, with and without contrast enhancement. A registration between this scanner and preoperative images (MRI or CT) was performed when available. For the delineation, the data were sent to the participating centers. The contours were next centralized. At the end of this first step, an interobserver analysis was performed to determine the IOV of flap contouring for each case. For the flaps for which the reproducibility was poor (Dice Similarity Coefficient, DSC < 0.7, see below), a second step was planned. For these flaps, tutorials were created to remind radiation oncologists of the nature of the flaps, their use and their aspect in imaging (Supplemental data). They were

Table 1. Patients' characteristics.

Patient	Location	Stage	Preoperative imaging	Surgery	Neck dissection	Flap
Step 1						
Pedicled flaps						
Patient A	Parotid	T4N1M0	Portal phase CE – CT	Parotidectomy	Ipsilateral	pectoralis
Patient B	Oral cavity (floor of the mouth)	T2bN0M0	Enhanced fat-suppressed T1-weighted MR	Tongue and floor resection	Bilateral	FAMM
Free flaps						
Patient C	Oral cavity (floor of the mouth)	T2N2bM0	portal phase CE – CT, first enhanced fat-suppressed T1-weighted MR	Tongue and floor resection	Bilateral	Radial forearm
Patient D	Oropharynx	T3N2bM0	portal phase CE – CT, first enhanced fat-suppressed T1-weighted MR	Oropharyngectomy	Bilateral	ALT
Patient E	Oral cavity (floor of the mouth)	T3N2bM0	portal phase CE – CT	Segmental mandibulectomy, tongue and floor resection	Bilateral	Fibula
Step 2						
Patient F	Oral cavity (floor of the mouth)	T1N2cM0	Enhanced fat-suppressed T1-weighted MR	Tongue and floor resection	Bilateral	FAMM

CE – CT: Contrast enhanced computed tomography; FAMM: facial artery musculo-mucosal flap; ALT: anterolateral thigh flap.

produced with the collaboration of surgeons and radiologists specializing in HNC surgery (VB, ALC, PYM). The images in these tutorials were extracted from unpublished images of the atlas. During the second step, each radiation oncologist of the working group was supposed to delineate without any look at its first contours, the case(s) which were not considered as reproducible during the first step. They must also delineate a new case of this flap, in order to evaluate the value of the tutorials and training for reducing IOV in the delineation of this flap.

Methods of comparison

The evaluation of the variability of contours was benchmarked using a 'reference' contour. The reference contour was created by a radiation oncologist who had participated in atlas creation [11], had more than 10 years' experience in contouring HNC patients after surgery, and belonged to one of the European centers with the most extensive experience in use of flaps (AC, from Department of Radiation Oncology, University Hospital of Amiens, Amiens. France). Radiation oncologists in the working group delineated each flap. LIFEx® software version 7.0.0 was used to evaluate and compare the contours, both qualitatively and quantitatively [12]. The volume (cc) for each delineation was computed and was compared to the volume of the 'reference' contour. Comparisons of the volumes were based on three indices: the volume ratio (VR) between the volume of each member of the working group (C_n) and the volume of the 'reference' contour (C_R), the common delineated volume (CDV $= (Cn \cap C_R)/C_R$) and the additional delineated volume (ADV $= (Cn - C_R)/C_R$). For assessing contour similarity [13], the Dice Similarity Coefficient (DSC) was calculated [13]. Contours were considered as reproducible if the median DSC > 0.7 (optimal value of DCS $= 1$). Indeed, several previous studies that evaluated IOV of OAR or target volume delineation reported that a DSC >0.7 may be sufficient to characterize the contours as reproducible [14–16]. This threshold is discussed in the discussion section. The coefficient of variation (ratio of the standard deviation/mean) of the volume measured between the different people of the WG for each FLAP was also computed with an optimal value in this case as low as possible. The other indices were analyzed with no threshold value but with the optimal values in mind: 100% for RV, 100% for CDV, and 0% for ADV.

Statistical analysis

Statistical analysis was performed with SAS software, version 9.4 (SAS Institute Inc., Cary, NC). The normality of the distribution for the volume and the four indexes was investigated with the Shapiro-Wilks test. Results were expressed as median and interquartile range for all parameters whatever the distribution. Difference between the volumes and the experts' volume was computed and compared to 0 with a student *t*-test or a Wilcoxon signed rank test. Comparative analysis of cases selected for blind recontouring between step 1 and step 2 was done with the Student *t*-test or Wilcoxon test for paired samples. Values of two-sided $p < 0.05$ were considered statistically significant. This study was approved by the Scientific Council of GORTEC on 27 August 2020.

Results

The flow chart on Supplemental Figure 1 summarizes the different steps of the study. Table 1 summarizes the main characteristics of the selected patients and Table 2 summarizes the main results obtained during interobserver analysis.

Similarity index in the first interobserver analysis

The median DSC obtained by the working group compared to the expert for the pectoralis flap, the FAMM flap, the radial forearm flap, ALT flap, and the fibular flap were 0.82, 0.40, 0.76, 0.81, and 0.76, respectively (Figure 1). For FAMM, this coefficient was 59.9 while for the other flaps of step 1, it was 39.6, 21.9, 19.8 and 29.7 for A – pectoralis, C – radial forearm, D – ALT and E – Fibula respectively. On this primary endpoint, delineation of the pectoralis, radial forearm, ALT and fibular flaps was considered reproducible and here was no need for a second step. For the FAMM flap, delineation was not considered reproducible and a second step was performed. Therefore, the patient B's FAMM flap was delineated again during the second step (Patient B – step 2) and a patient F, with a new FAMM flap, was also evaluated.

Volume comparison in the first interobserver analysis

In the first analysis (step 1), the median volumes of the five flaps delineated by the working group were close to the expert contour value, with approximately 50 cc for the pectoral, fibula, and ALT flaps, 20 cc for the radial forearm, and up to 10 cc for the FAMM flap (Table 2). The volumetric ratio was thus close to the optimal value of 100% for all flaps. Moreover, the CDV and the ADV were close to their respective optimal value of 100% (84–90%) and 0% (22–31%), except for the FAMM flap for which the CDV and the ADV were 53% and 52% respectively.

Similarity index and volume comparison of FAMM delineation in the second interobserver analysis

The DSC obtained by the working group compared to the expert for the same FAMM flap (patient B – step 2) and a new FAMM flap (patient F) were 0.54 and 0.33, respectively. In this step 2, the volume of both FAMM flaps delineated by the expert were 8.97 (patient B – step 2) and 9.06 cc (patient F), respectively. The coefficient of variation was 51.7 for B – FAMM and 46.2 for F – FAMM. The median volume of both FAMM flaps delineated by the working group were 6.5 (patient B – step 2) and 7.6 cc (patient F), respectively. The volume indices obtained by the working group compared to the expert for the same FAMM flap (patient B – step 2) were 59.2% for the volume ratio, 30.9% for the CDV, 49.9% for the ADV (Figure 2). The different indices obtained by the

Table 2. Similarity index description and volume comparison for each type of flap and at each step.

	Patient – FLAP	Step 1	Step 2	p-Value (step 1 vs step 2)
Dice Similarity Coefficient (DSC)[a]	A – pectoralis	0.82 (0.42–0.87)	–	–
DSC = 2× (Cn ∩ CR)/(Cn + CR)	B – FAMM	0.40 (0.37–0.46)	0.54 (0.42–0.56)	0.124
	C – radial forearm	0.76 (0.74–0.79)	–	–
Optimal value = 1				
	D – ALT	0.81 (0.79–0.87)	–	–
	E – Fibula	0.76 (0.71–0.78)	–	–
	F – FAMM	–	0.34 (0.23–0.49)	–
Volume (cc)[b]	A – pectoralis	59.43 (54.72–173.89) / 51.99 (p = 0.078)	–	–
	B – FAMM	3.55 (3.47–8.92) / 8.87 (p = 0.36)	7.06 (4.24–9.04)/ 8.97 (p = 0.22)	0.131
	C – radial forearm	22.6 (18.38–23.78)/ 21.21 (p = 0.64)	–	–
	D – ALT	55.25 (44.66–70.9) / 47.81 (p = 0.15)	–	–
	E – Fibula	63.3 (39.61–69.8) / 51.09 (p = 0.31)	–	–
	F – FAMM	–	7.6 (5.34–10.17) / 9.06 (p = 0.48)	–
Volume Ratio (VR)[a]	A – pectoralis	114.3 (105.03–334.45)	–	–
VR = Cn/CR	B – FAMM	100.59 (39.15–170.92)	59.17 (47.8–101.95)	0.583
Optimal value = 100%	C – radial forearm	89.57 (86.63–112.11)	–	–
	D – ALT	115.55 (93.41–148.29)	–	–
	E – Fibula	123.89 (77.54–136.63)	–	–
	F – FAMM	–	80.19 (59–112.28)	–
Common Delineated Volume (CDV)[a]	A – pectoralis	89.95 (85.54–92.49)	–	–
CDV = (Cn ∩ CR)/CR	B – FAMM	53.42 (30.82–65.63)	30.97 (30.82–37.56)	0.227
Optimal value = 100%	C – radial forearm	75.03 (72.35–83.55)	–	–
	D – ALT	89.34 (81.21–97.65)	–	–
	E – Fibula	84.47 (69.51–91.12)	–	–
	F – FAMM		33.32 (18.37–41.59)	–
Additional Delineated Volume (ADV)[a]	A – pectoralis	22.45 (13.46–72.25)	–	–
ADV = (Cn – CR)/CR	B – FAMM	52.46 (34.38–57.3)	49.92 (21.43–55.99)	0.877
Optimal value = 0%	C – radial forearm	25.47 (17.24–38.22)	–	–
	D – ALT	22.28 (10.79–34.15)	–	–
	E – Fibula	31.26 (13.1–34.58)	–	–
	F – FAMM	–	47.69 (40.35–80.28)	–

[a]Median and interquartile range for the Working group.
[b]Median and interquartile range for the Working group/Expert contour (cc), (p-value of the difference).
FAMM: facial artery musculo-mucosal flap; ALT: anterolateral thigh flap.

working group compared to the expert for a new FAMM flap (patient F) also showed poor interobserver reproducibility with 80.1% for the volume ratio, 32.3% for the CDV, and 47.7% for the ADV (Table 2).

Discussion

This study evaluated IOV in the volume delineation of flaps used in HNC reconstructive surgery and proposed tutorials that complemented the published atlas [11] to reduce interobserver variability. It is important to note that the imaging tutorials were derived from data not published in the atlas.

In the present study, we selected five different flaps. These are the ones most frequently used in reconstructive surgery for HNC [6], whose delineation has been described in the atlas [11]. For four of these flaps (the regional pedicled pectoralis major myocutaneous flap, the fasciocutaneous radial forearm free flap, the soft tissue anterolateral thigh free flap, and the fibular free flap), the Dice Similarity Coefficient (DSC) was $\geq$0.70.

To our knowledge, no study has yet evaluated the IOV of target volume delineation for HN surgically treated cancers with a flap. In 2017, Christiaens et al. performed a benchmark case procedure before patient inclusion in the phase III EORTC 1219-DAHANCA 29 intergroup trial [14]. The participating centers were asked to perform a 2-step benchmark case including a delineation exercise according to the protocol guidelines. The mean DSC for the accepted high-risk CTV was 0.75. In this trial, patients had not been operated on for their locally advanced HNC. The definition of target volumes is more evident for these non-operated patients, with well-established guidelines [4,17], than for operated patients, with few guidelines [5]. Besides, in our study, some of the differences observed were often due to a misunderstanding of the instructions (some participants had contoured the CTV and included the surgical clip for patient with the radial forearm), or to a misreading of the surgical report (in particular for the patient with fibula flap, for which some did not include the bone part of the flap).

Besides this, IOV in OAR delineation was not a requirement for admission to the trial in the EORTC 1219-DAHANCA 29 intergroup trial but was nevertheless assessed [14]. Mean DSC appeared to be 0.45, 0.5, and 0.8 for the larynx, the oral cavity, and the parotids, respectively. In another study, Brouwer et al. evaluated the IOV for five head and neck OAR: spinal cord, parotid and submandibular glands, thyroid cartilage, and glottic larynx [15]. The mean concordance index values varied from 0.64 to 0.71, except for the glottic larynx for which the mean CI was 0.37. In another recent study, Van der Veen et al. [16] also evaluated IOV in OAR delineation in HNC, based on the international guidelines published in 2015 [18]. For OARs that are clearly identified on the scanner and relatively large (brain stem, mandible, salivary glands),

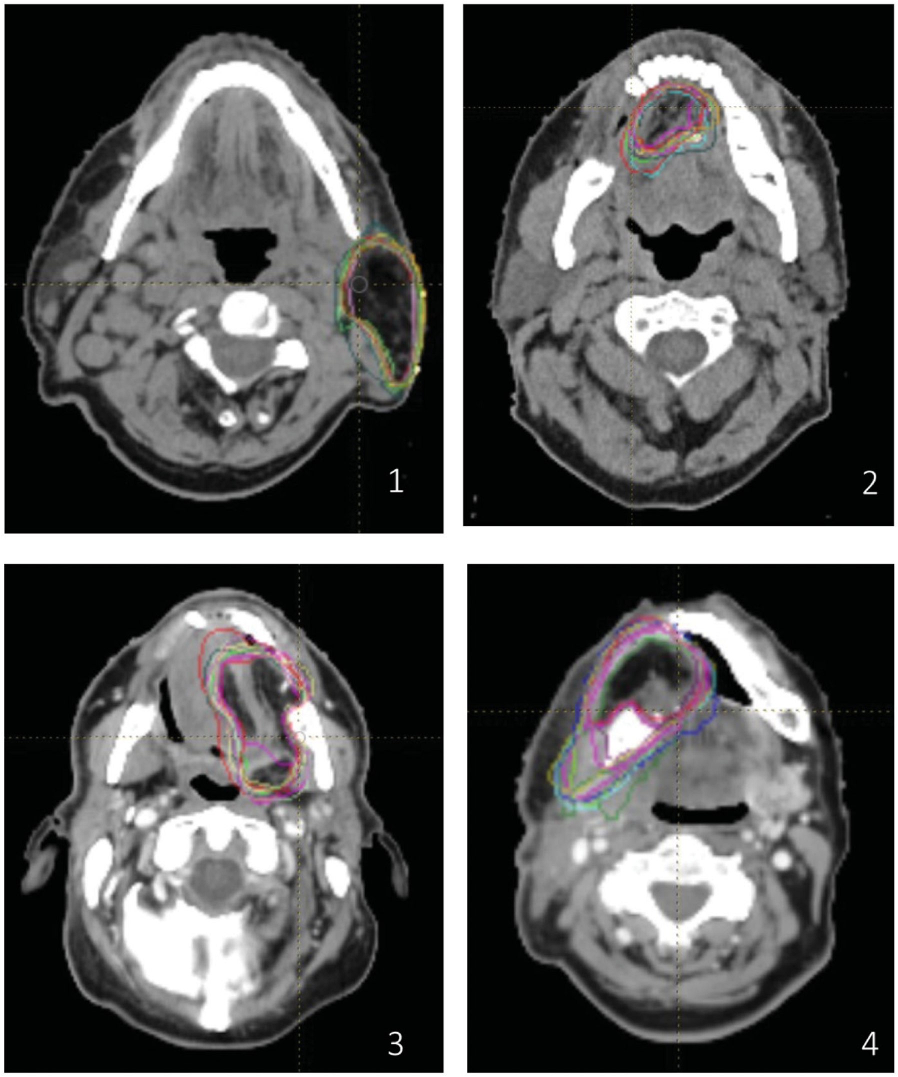

Figure 1. First inter-observer analysis for flap delineation. This figure represents the results obtained for the first inter-observer analysis (step 1) for (1) the pectoralis major myocutaneous flap (patient A), (2) the radial forearm flap (patient C), (3) the anterolateral thigh flap (patient D), and (4) the fibular flap (patient E). This analysis showed overall that the delineation of these flaps was reproducible (with a Dice Similarity Coefficient > 0.7).

the DSC was high (>0.8). For smaller or less clearly identifiable organs (oral cavity, spinal cord, larynx), the DSC was close to 0.7. For small organs, such as the cochlea, the DSC was much lower (DSC < 0.5). Thus, whether as a target volume, or OAR, for head and neck cancer, a DSC > 0.7 was therefore usually considered sufficient to characterize the contours as reproducible.

Therefore, for the larger musculocutaneous or bone flaps, i.e. the regional pedicled pectoralis major myocutaneous flap, the soft tissue anterolateral thigh free flap, the fibular free flap, and the medium size fasciocutaneous radial forearm free flap, the delineation was considered as reproducible from the first step and was not retested during the second step.

Unlike the previous four flaps, delineation of the local FAMM flap had very poor reproducibility at the time of the first interobserver analysis. There was a slight improvement during step 2 for the case delineated a second time, indicating the value of the training and the tutorial, when using additional thin slices of the published atlas, although it remained poorly reproducible overall.

The lack of reproducibility for FAMM flap delineation could be due to the complex visualization of the flap on postoperative imaging, including CT [19]. A postoperative MRI could possibly help the radiation oncologist to delineate the FAMM flap, especially the T1 and T2-weighted MR, and the CUBE enhanced fat-suppressed T1-weighted MR. This was however beyond the scope of the current study. It is important to remember that this flap is usually short, harvested in a plane deep to the facial artery by including the overlying part of the buccinator muscle along its length and part of the orbicularis oris in the area of the oral commissure [20]. The FAMM flap is rotated next to its native mucosal cheek area. It does not add any unusual tissue such as bone, thick muscle, or artificial material in the reconstructed area, which may have made it easier to identify on imaging. Therefore, even more than for other flaps, the operative report is necessary to accurately locate the flap site and its

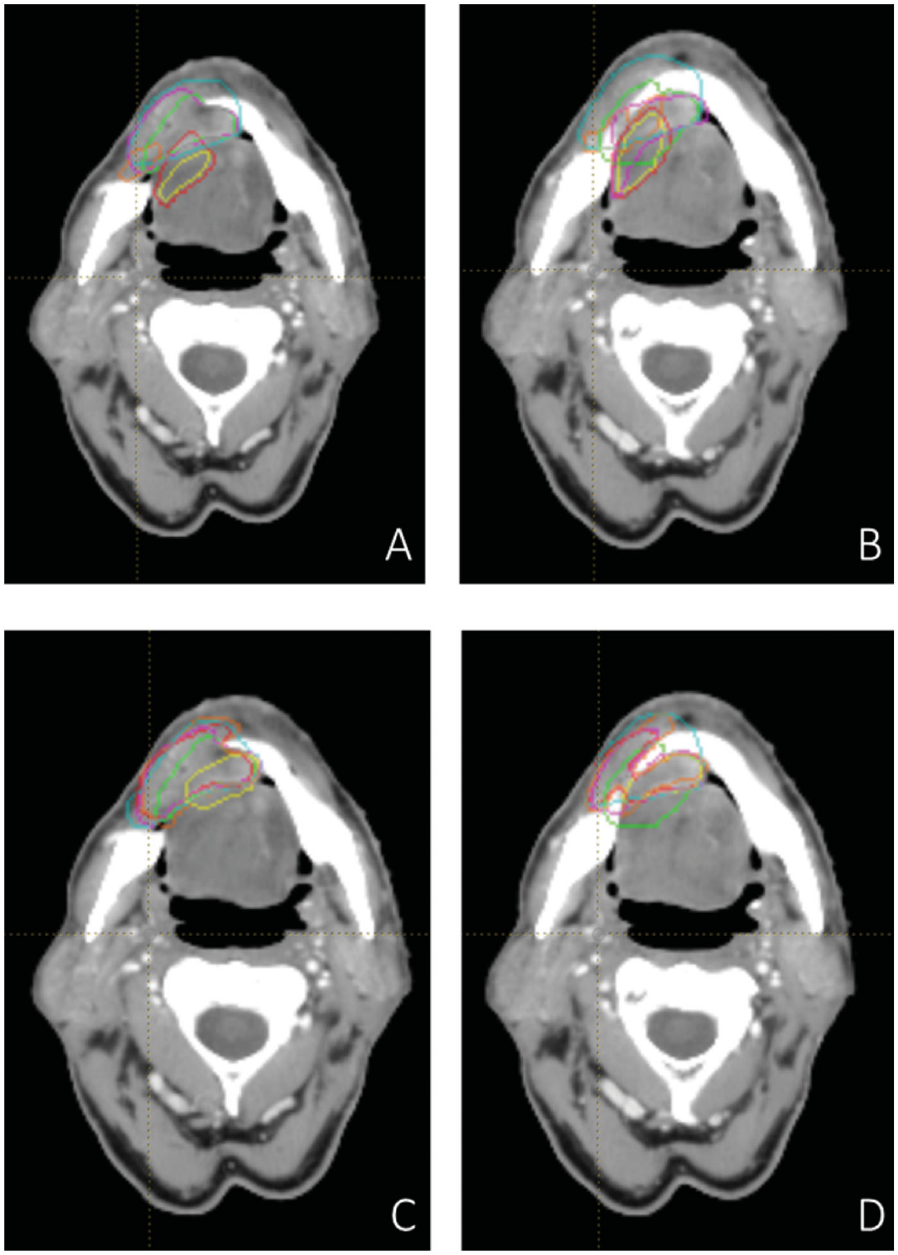

Figure 2. Comparison between the first and second inter-observer analysis for the FAMM. This figure shows the workgroup contours for patient B (FAMM). Images A and B correspond to the first analysis, images C and D correspond to the second. Visually, it appeared that the contours were more reproducible in the second step but the quantitative analyses just found a slightly improvement.

components based on their tissue densities, length and thickness after flap harvesting and reshaping. A standardized operative report should include precise information, already listed in the atlas [11].

It is also important to note that the DSC could be lower for the FAMM flap because it was the smallest structure. As in the studies mentioned above, our results were based on DSC which is the one most often used in interobserver studies: it is simple to use and understand and allows to appreciate the intersection between two contours. However, the DSC could tend to be lower for small structures: the shift of one or two pixels on a small volume (of a few pixels) may result in a major impact whereas this same shift could be negligible on volumes of a few hundred pixels. Since we are unable to calculate the widely used Hausdorff distance with our software, our results according to DSC were supported by calculating the coefficient of variation of the volume, another measure that is independent of the size of the volume. This coefficient was overall higher for the FAMM than for the other flaps.

Besides this, an expansion around small structures (for which interobserver variability is large) on RT quality images could provide more robust (not more precise) documentation of dose volumes effects. Therefore, it might also be useful, to assess flap outcomes, to add an isocentric margin of 5 mm, to consider delineation heterogeneities and delineation uncertainty OAR volume to assess dose volume relationships.

Finally, it is interesting to note that the FAMM flap is not always popular in surgical teams specializing in maxillofacial

Table 3. Possible post-operative radiation-induced toxicity after HNC reconstructive surgery.

	Number of patients	Cancer localization	Type of Flap (Majority)	Toxicity
Bozec et al. 2009 [21]	132 (86 RT)	Oral cavity and oropharynx	Radial forearm free flap	Speech, Mouth opening Esthetic outcome
Shin et al. 2012 [9]	31	Oral cavity (tongue)	Radial forearm free flap	Speech outcomes
Pierre et al. 2014 [22]	64 (48 RT)	Oral cavity and oropharynx	Radial forearm free flap and fibular free flap	Nasal speech, swallowing, pharyngeal retention, mouth opening
Van Hinte et al. 2019 [23]	113 (57)	Oral cavity	Myocutaneous free flap	Neck and shoulder function
Petrovic et al. 2019 [24]	25	Oral cavity	fibular free flap	Esthetic outcome
Yi et al. 2020 [25]	30	Oral cavity (tongue)	Radial forearm free flap and anterolateral thigh flap	Speech outcomes
Dziegielewski et al. 2020	74	HNC (not specified)	Bone flaps (especially fibular free flap)	Osteoradionecrosis

reconstruction because of limited versatility (small size, mucosa only), rotation constraints, and it has been associated with scar bridges [20]. For these reasons, free flaps are more often used, especially the antebrachial for medium defects or the ALT for large defects. While assessment of flap outcomes should include FAMMs, FAMM-volume sparing using IMRT may also be less relevant and possibly at risk for tumor failure as this flap is very thin.

There are currently no specific recommendations for CTV delineation in the postoperative setting. As shown in the recent delphi flap study [8], some teams tend to include the entire flap in the CTV while others include only a part of the flap. The inclusion of the entire flap could have consequences on toxicities. Several surgical reports have suggested that RT had deleterious effects on flaps with respect to functional outcomes (Table 3). In a recent study, Yi et al. retrospectively analyzed outcomes of patients who underwent hemi-tongue reconstruction followed by speech therapy for 4 years. Patients who had postoperative radiotherapy (PORT) had a lower speech outcome and precise measurements showed that the contralateralization of the tongue was affected by RT [26]. However, the authors did not find a correlation between the prescribed dose and the observed toxicities. Shin et al. also precisely studied speech integrity after tongue resection and PORT [9]. Tongue mobility was clearly affected by PORT. In two others studies evaluating long-term functional outcomes and quality of life after oncologic surgery and microvascular reconstruction in patients with oral or oropharyngeal cancer, PORT was the main predictive factor of worse functional outcome included speech, mouth opening, swallowing, pharyngeal retention, and esthetic outcomes [21,25]. In a recent study, Pterovic et al. showed that at least a part of these side effects could be due to radiation-induced soft and bone tissue atrophy [22]. In another study, Van Hinte et al. also showed that patients treated with reconstructive surgery for oral cavity cancer sometimes extending to the mandible, with different types of flaps, had impaired shoulder function after PORT [23]. Nevertheless, in this study, the patients receiving postoperative RT had significantly more extensive surgery than patients undergoing surgery alone ($p = 0.023$). In addition, no dose-volume information was reported, only dose range 54–70 Gy. Moreover, in all these studies, the authors did not specify if the flaps and pedicles were included in the CTV. The results reported in Table 3 are therefore imprecise and possibly impaired by the lack of data on the real dose received by the flaps. The delineation of the postoperative CTV after reconstructive surgery for HNC is complicated because it is not known if the flap should be included in its entirety in the CTV, or if on the contrary it should be protected [8]. Radiation oncologists are naturally inclined to view any abnormal tissue as a ‘target’. Residents and non-head and neck radiation oncology specialists may even contour the body of a head and neck flap as a high-risk tumor volume, while completely missing the recipient bed. Indeed, in a recent study, Cho et al. evaluated the local recurrence patterns of 114 patients with HNC treated with curative resection and reconstruction for oropharyngeal and oral cancers followed by PORT [24]. None of the local relapses was located in the flap tissue, whereas the majority were ≤ 5 mm from the flap anastomosis. These results emphasize the need for clear recommendations and interobserver reproducibility for flaps delineation, in order to precisely evaluate the impact of flaps irradiation and define clear recommendations for CTV delineation in PORT setting.

This is the first interobserver study of flap delineation for head and neck cancers. An important strength of this study is that it is multicenter and international, as the 14 radiation oncologists in the working group belong to centers worldwide. Notably, in some of the participating countries, surgery is rarely performed in intermediate – locally advanced stages and postoperative radiotherapy is rare. This study therefore integrates the differences in practice in these different countries, and thus the significant result obtained for four of the main flaps currently used in reconstructive surgery for head and neck cancers is likely truly reproducible. It should be noted that, as much as possible, the delineation of these complex structures, whatever their size and location, should be performed in coordination with the surgeon and the specialized radiologist. Nevertheless, DSC > 0.7 seemed reasonable considering the purpose of the X flap program: to document the pattern of failures and toxicities of patients with flaps, without changing the standards in terms of definitions and dose volume constraints. However, for delineation of the FAMM flap, particular care is needed whereby the radiation oncologists should use surgical and imaging reports and work hand in hand with the surgeon and radiologist to try to define these structures as well as possible. This study is an important starting point for any study assessing post-operative radiotherapy in the presence of flaps. It provides the basis for continuing study of the precise impact of including the whole flap in the target volume, both for local control and the risk of toxicities in operated HNC.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Sue S. Yom http://orcid.org/0000-0002-0779-7476
Vianney Bastit http://orcid.org/0000-0001-7394-1301

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

References

[1] Mishra RC, Singh DN, Mishra TK. Post-operative radiotherapy in carcinoma of buccal mucosa, a prospective randomized trial. Eur J Surg Oncol. 1996;22(5):502–504.
[2] Al-Sarraf M, Pajak TF, Byhardt RW, et al. Postoperative radiotherapy with concurrent cisplatin appears to improve locoregional control of advanced, resectable head and neck cancers: RTOG 88-24. Int J Radiat Oncol Biol Phys. 1997;37(4):777–782.
[3] Hosni A, Huang SH, Chiu K, et al. Predictors of early recurrence prior to planned postoperative radiation therapy for oral cavity squamous cell carcinoma and outcomes following salvage intensified radiation therapy. Int J Radiat Oncol*Biol*Phys. 2019;103(2):363–373.
[4] Grégoire V, Evans M, Le Q-T, et al. Delineation of the primary tumour Clinical Target Volumes (CTV-P) in laryngeal, hypopharyngeal, oropharyngeal and oral cavity squamous cell carcinoma: AIRO, CACA, DAHANCA, EORTC, GEORCC, GORTEC, HKNPCSG, HNCIG, IAG-KHT, LPRHHT, NCIC CTG, NCRI, NRG oncology, PHNS, SBRT, SOMERA, SRO, SSHNO, TROG consensus guidelines. Radiother Oncol. 2018;126(1):3–24.
[5] Evans M, Beasley M. Target delineation for postoperative treatment of head and neck cancer. Oral Oncol. 2018;86:288–295.
[6] Urken ML. Advances in head and neck reconstruction. Laryngoscope. 2003;113(9):1473–1476.
[7] Thariat J, Leleu T, Micault E, et al. Dix ans de progrès en chirurgie des tumeurs de la tête et du cou, quel impact sur la radiothérapie postopératoire ? Bulletin du Cancer. 2020;107(7–8):823–829.
[8] Carsuzaa F, Lapeyre M, Gregoire V, et al. Recommendations for postoperative radiotherapy in head & neck squamous cell carcinoma in the presence of flaps: a GORTEC internationally-reviewed HNCIG-endorsed consensus. Radiother Oncol. 2021;160:140–147.
[9] Shin YS, Koh YW, Kim S-H, et al. Radiotherapy deteriorates postoperative functional outcome after partial glossectomy with free flap reconstruction. J Oral Maxillofac Surg. 2012;70(1):216–220.
[10] Winter SCA, Cassell O, Corbridge RJ, et al. Quality of life following resection, free flap reconstruction and postoperative external beam radiotherapy for squamous cell carcinoma of the base of tongue. Clin Otolaryngol Allied Sci. 2004;29(3):274–278.
[11] Le Guevelou J, Bastit V, Marcy PY, et al. Flap delineation guidelines in postoperative head and neck radiation therapy for head and neck cancers. Radiother Oncol. 2020;151:256–265.
[12] Nioche C, Orlhac F, Boughdad S, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. Cancer Res. 2018;78(16):4786–4789.
[13] Jameson MG, Holloway LC, Vial PJ, et al. A review of methods of analysis in contouring studies for radiation oncology: techniques of contour comparison. J Med Imaging Radiat Oncol. 2010;54(5):401–410.
[14] Christiaens M, Collette S, Overgaard J, et al. Quality assurance of radiotherapy in the ongoing EORTC 1219-DAHANCA-29 trial for HPV/p16 negative squamous cell carcinoma of the head and neck: results of the benchmark case procedure. Radiother Oncol. 2017;123(3):424–430.
[15] Brouwer CL, Steenbakkers RJ, van den Heuvel E, et al. 3D variation in delineation of head and neck organs at risk. Radiat Oncol. 2012;7:32.
[16] van der Veen J, Gulyban A, Willems S, et al. Interobserver variability in organ at risk delineation in head and neck cancer. Radiat Oncol. 2021;16(1):120.
[17] Racadot S, Vérillaud B, Serre A-A, et al. Impact of reconstructive or minimal invasive surgery on the assessment of current definitions of postoperative clinical target volume for head and neck cancers. Cancer Radiother. 2020;24(6–7):649–657.
[18] Brouwer CL, Steenbakkers RJHM, Bourhis J, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG oncology and TROG consensus guidelines. Radiother Oncol. 2015;117(1):83–90.
[19] Syed F, Spector ME, Cornelius R, et al. Head and neck reconstructive surgery: what the radiologist needs to know. Eur Radiol. 2016;26(10):3345–3352.
[20] Ayad T, Xie L. Facial artery musculomucosal flap in head and neck reconstruction: a systematic review. Head Neck. 2015;37(9):1375–1386.
[21] Pierre CS, Dassonville O, Chamorey E, et al. Long-term functional outcomes and quality of life after oncologic surgery and microvascular reconstruction in patients with oral or oropharyngeal cancer. Acta Otolaryngol. 2014;134(10):1086–1093.
[22] Petrovic I, Baser R, Blackwell T, et al. Long term functional and esthetic outcomes after fibula free flap reconstruction of the mandible. Head Neck. 2019;41(7):2123–2132.
[23] van Hinte G, Wetzels J-W, Merkx MAW, et al. Factors influencing neck and shoulder function after oral oncology treatment: a five-year prospective cohort study in 113 patients. Support Care Cancer. 2019;27(7):2553–2560.
[24] Cho Y, Yoon HI, Lee IJ, et al. Patterns of local recurrence after curative resection and reconstruction for oropharyngeal and oral cancers: Implications for postoperative radiotherapy target volumes. Head Neck. 2019;41(11):3916–3923.
[25] Bozec A, Poissonnet G, Chamorey E, et al. Radical ablative surgery and radial forearm free flap (RFFF) reconstruction for patients with oral or oropharyngeal cancer: postoperative outcomes and oncologic and functional results. Acta Otolaryngol. 2009;129(6):681–687.
[26] Yi CR, Jeong WS, Oh TS, et al. Analysis of speech and functional outcomes in tongue reconstruction after hemiglossectomy. J Reconstr Microsurg. 2020;36(07):507–513.