Review paper

Practical and technical key challenges in head and neck adaptive radiotherapy: The GORTEC point of view

Nolwenn Delaby a, Anaïs Barateau b, Sophie Chiavassa c, Marie-Claude Biston d,e, Philippe Chartier f, Eliane Graulüres g, Lucie Guinément b, Sandrine Huger b, Thomas Lacornerie i, Corinne Millardet-Martin l, Alain Sottiaux k, Jérôme Caron l, David Gensanne m,n, Yoann Pointreau o, Alexandre Coutte p, Julian Blau q,r, Anne-Agathe Serre d, Joel Castelli b, Milan Tomsej k, Robin Garcia s, Catherine Khamphan s, Aurélien Badey s, *

a Centre Eugène Marquis, Unité de Physique Médicale, rue de La Bataille Flandres Dunkerque - CS 44229, 59042 Rennes Cedex, France
b Univ Rennes, CLCC Eugène Marquis, Inserm, LTSI – UMR 1099, 35000 Rennes, France
c Institut de cancérologie de l’Ouest (ICO) Centre René-Gauducheau, Service de physique médicale, 44805 Saint-Herbain, France.
d Université de Nantes, CHU de Nantes, 15, avenue des Buttes de Coësmes, 44093 Nantes, France
e Centre Henri Becquerel, 69373, LYON Cedex 08
f Centre Hospitalier Universitaire Grenoble Alpes, Département de Radiothérapie, BP217, 38043 Grenoble, Cedex 9, France
g CENTRAL, CNRS UMR5220, Inserm U1044, INSIA-Lyon, Université Lyon 1, 69100 Villeurbanne, France
h Centre Hospitalier Universitaire de Lyon, Département de Physique Médicale, 69287 Lyon 09, France
i Centre Oscar Lambret, Service de Physique Médicale, 59000 Lille, France
j Centre Jean Perrin, Département de Physique Médicale, 58, rue Montalembert, BP 392, 63011 Clermont-Ferrand Cedex 1, France
k CHU de Charleroi, Département de Radiologie, B-6110 Charleroi, Belgique
l Institut Bergonie, Département de radiothérapie, 33000 Bordeaux, France
m Centre Henri Becquerel, Département de radiothérapie, 76000 Rouen, France
n Centre Henri Becquerel, Département de radiothérapie, 76000 Rouen, France
o QUANTIFI-LITIS EA4108, Université de Rouen Normandie, Rouen, France
p Centre Louis Chauvel, Institute of Cancerology, Centre de Cancérologie de la Sarthe (CCS), BP 392, 72000 Le Mans, France
q CHU Amiens Picardie, Département de radiothérapie, 80054 Amiens, France
r Centre Jean Perrin, Département de Radiothérapie, 58, rue Montalembert, BP 392, 63011 Clermont-Ferrand Cedex 1, France
s INERIM U1240 IMoST, Université de Clermont Auvergne, 63000 Clermont-Ferrand, France
t Centre Louis Gauducheau, Service de physique médicale, 84918 Avignon Cedex 9, France

* Corresponding author.
E-mail address: abadey@iuc84.org (A. Badey).

https://doi.org/10.1016/j.ejmp.2023.102568
Received 13 July 2022; Received in revised form 15 February 2023; Accepted 18 March 2023
Available online 2 April 2023

Contents lists available at ScienceDirect
Physica Medica
journal homepage: www.elsevier.com/locate/ejmp

ARTICLE INFO

Keywords:
Adaptive radiotherapy
Head and neck cancer
Automatic segmentation
Image registration
Dose accumulation
Quality assurance
Workflow

ABSTRACT

Anatomical variations occur during head and neck (H&N) radiotherapy (RT) treatment. These variations may result in underdosage to the target volume or overdosage to the organ at risk. Replanning during the treatment course can be triggered to overcome this issue. Due to technological, methodological and clinical evolutions, tools for adaptive RT (ART) are becoming increasingly sophisticated. The aim of this paper is to give an overview of the key steps of an H&N ART workflow and tools from the point of view of a group of French-speaking medical physicists and physicians (from GORTEC). Focuses are made on image registration, segmentation, estimation of the delivered dose of the day, workflow and quality assurance for an implementation of H&N offline and online ART. Practical recommendations are given to assist physicians and medical physicists in a clinical workflow.

1. Introduction

Developments in intensity modulated (IMRT) and image-guided radiotherapy (IGRT) devices have allowed more precise and targeted head and neck (H&N) cancer treatments, with improved sparing of organs at risk (OARs) while covering target volumes. However, during H&N RT, some patients are typically subject to anatomical variations such as tumor shrinkage or weight loss. These changes are progressive...
ongoing and future multi-centre clinical trials must be robust to prove clinical benefit has not yet been formally demonstrated for all patients and/or dosimetric variations observed during RT [5,7]. However, the use of offline adaptation [6] or for studies evaluating anatomical strategies are currently being evaluated and tested in clinical trials to treat treatment in relation to functional imaging is strongly correlated with allows an efficient deployment in RT departments. These new features (especially for online strategies) imply the need to set up evaluations and rigour in the implementation to treatment delivery. These new features (especially for online strategies) imply the need to set up evaluations and rigour in the implementation to allow an efficient deployment in RT departments.

ART provides a dosimetric benefit for a majority of clinical studies that used offline adaptation [6] or for studies evaluating anatomical and/or dosimetric variations observed during RT [5,7]. However, the clinical benefit has not yet been formally demonstrated for all patients [7] but rather for specific cohorts of patients [17]. For these reasons, ongoing and future multi-centre clinical trials must be robust to prove the effectiveness of H&N ART strategies. However, practices and commercial tools can differ between participating centers. Thus, recommendations and/or quality control need to be set to homogenise practices. The implementation of ART strategies may induce additional uncertainties through new steps implemented during preparation and treatment delivery. These new features (especially for online strategies) imply the need to set up evaluations and rigour in the implementation to allow an efficient deployment in RT departments.

A group of 26 medical physicists together with a group of 8 radiation oncologists have discussed the implementation of H&N ART within the GORTEC (Radiotherapy Oncology Group for Head and Neck). Created in 1999, the GORTEC involves 150 oncologists and 30 medical physicists. This network covers more than 100 oncology care institutions established in France, Switzerland, Belgium, Tunisia, Germany, and Spain. The aim of this paper is to give an overview of the key steps of an H&N ART workflow and tools from the point of view of a group of French-speaking medical physicists and physicians from GORTEC, to provide practical considerations for the implementation of an H&N ART process. Focuses were made on image registration, segmentation, estimation of the delivered dose of the day, dose monitoring, offline and online workflows and quality assurance for an implementation of H&N ART.

2. Image registration

Image registration is a key step in an ART workflow. The registration of two series of images consists, through geometric transformations, in matching their intrinsic elements. For RT, rigid image registration (RIR) is used typically to position the patient by registering images acquired each day to the planning kV-CT. Deformable image registration (DIR) can be performed for contour propagation, planning adaptation or dose accumulation during the treatment course [18,19].

2.1. Rigid image registration (RIR)

In RT, RIR can be used at different steps of the patient’s treatment, in particular to position the patient before irradiation. RIR is a transformation preserving the distance between all points in the image. RIR can be performed only with translations, or with translations and rotations, in 2D or 3D. RIR is subject to uncertainties evaluated with physical or numerical phantoms. Since RIR uncertainties depend on imaging modality, ideal physical phantoms would be compatible with all imaging modalities. Geometrical phantoms can be used to assess spatial integrity of RIR, and anthropomorphic phantoms to evaluate clinical RIR uncertainties. AAPM Task Group 132 (TG-132) provides data to assess RIR uncertainties with both geometrical and anthropomorphic phantoms for several imaging modalities (kV-CT, kV-CBCT, PET, MRI) [20]. An evaluation of RIR is recommended before the first software use, and at each update.

2.2. Deformable image registration (DIR)

DIR is an essential tool used during the ART process to take into account and quantify changes in the shape and size of internal organs to adapt between initial planning images and daily images acquired during the treatment course. The transformation between two image series consists in applying a deformation matrix to the image to be registered with the target: each pixel of the reference image finds its corresponding pixel in the image to be registered. This produces a deformation vector field (DVF) which is the mathematical vector translation of the change between two sets of images. Some solutions propose to visualize DVF mapping either by static or dynamic representation. In an evaluation process, it is usually possible to export the DVF generated by a DIR algorithm in DICOM format.

Traditionally, there are two main approaches in terms of functionality: feature-based and intensity-based methods. A third approach, called hybrid methods, combines the two previous ones [21,22]. DIR tools are available in many TPS or commercial softwares with different algorithms. DIR algorithms are used on a case-by-case basis and according to the center background for offline strategies. For online ART, DIR is used for contour propagation, synthetic CT generation and dose accumulation.

DIR performance is notably related to the input data, i.e. image quality of data to register, but also to DIR algorithm properties, the potential imprecision encountered and anatomical deformation magnitude observed. In case of vector fields smoothing use, it could be problematic for important anatomical changes. Artifacts and/or severe deformations between two image series will directly condition DIR quality. Indeed, larger errors were mostly encountered in regions around major shape changes, as well as areas with uniform contrast but large local motion discontinuity [23]. Geometrical parameters like matrix size and z-slice thickness of image series to register have an impact on the integrity and size of the voxels (resolution, partial volume effect, sampling, etc.) [20]. It is therefore important to assess the image quality beforehand and to work on image-guided (ART) protocol optimisation, particularly during acquisition and reconstruction. Moreover, DIR has several limitations: areas of artifacts, huge deformations, low contrast areas, sliding between tissues, matter appearance/disappearance, tumor shrinkage or multi-modal registration [24]. The uncertainty quantification associated with DIR algorithms is the subject of many works [25,26]. Indeed, although the algorithms are based on complex mathematical models, they are not consistent with biological/physiological processes.

For DIR evaluation, visual inspection is the first and simplest check that must be performed with split screen displays or regions of interest. Although it implies that the operator has the background to analyse this
type of transformation. AAPM Task Group 132 (TG-132) have provided recommendations on qualitative and quantitative metrics for evaluating the DIR process [20]. Numerical, geometrical, or anthropomorphic phantoms can be created and modified using dedicated software. They can mimic complex deformations observed in clinical practice. The deformed images can be compared to initial images. Filters and/or noise can be applied to approximate clinical conditions [27]. Deformation Vector Field (DVF) can be used to assess DIR. Groundtruth data is used to evaluate the performance of DIR algorithms by comparison with DVF obtained, to determine error at each point of the image. Some authors provide access to this data for evaluation of their own DIR algorithm [28,29]. Qualitative methods include for example an operator visualisation, such as comparing images with structure overlays. Quantitative verification is one of the challenges of DIR. It is often not possible because of the lack of groundtruth. For contour propagation, when access to DVF analysis is not possible, the evaluation tools available to validate DIR solutions are: physical phantoms [30,31] and numerical phantoms [28] as well as patient images [29]. Quantitative methods can require patient delineations or anatomical markers. For contouring, the deformed contours are compared with so-called reference contours, generally produced by an expert operator. Quantitative metrics such as DICE index, Jacobian matrix, target registration error (TRE), mean distance to agreement (MDA), and consistency can be used [20]. None of these metrics are perfect, so a combination of several indexes, at least overlapping and distance evaluation metrics with complementary specificities, is recommended. TG-132 also provided recommendations on tolerances, based on the application and image voxel dimensions [20]. In addition, subjective scoring methods for evaluating the mapped structures have been proposed [32].

**Practical recommendations:**

- An evaluation of RIR and DIR is recommended before the first software use, and at each update.
- Geometrical phantoms must be used to assess spatial integrity of RIR, and anthropomorphic phantoms to evaluate clinical DIR uncertainties.
- Create known deformations from physical, numerical phantoms or patient images is recommended to evaluate initial DIR algorithm performance.
- In case of automatic registration (RIR or DIR), it is recommended to perform it in a localised area of interest (with boxes) and where information from the two images are available.
- A patient-specific DIR review is required (at least visual, quantitative if possible). Systematic visual inspection is necessary to validate deformation and to identify areas of uncertainty such as large deformations or artifacts.
- For DIR evaluation with contours, a combination of several indexes, at least overlapping and distance evaluation metrics with complementary specificities, is recommended.

3. Image segmentation for delineation

The definition of contours is a key stage since it conditions the entire treatment management process. First of all, in order to control the management of many organs to be delineated, it is required to use lists of names and descriptions rigorously drawn up with a view to standardisation. Several American and European groups have proposed a standard list of names for target volumes and OARs in RT [33–36]. Such standardisation may be necessary to ensure quality control of clinical trials.

H&N segmentation can be performed on different imaging modalities: kV-CT [37], MR [38] or PET [39], but also with in-room imaging modalities such as kV-CBCT [40] or MV-CT [41,42] or kV-CT [43]. MR segmentation for H&N localisation is even more investigated with MR-Linac deployment in the radiotherapy departments [44,45] and MR simulation [46].

Manual segmentation is time consuming and implies inter-operator variability, regardless of the image type [47] and tumor localisation [48]. So, initial contouring, offline and online ART processes need segmentation automation [49].

Automatic segmentation (AS) for delineation, provided from initial planning and in-room imaging, allows to obtain new contours for example for evaluation of anatomical and/or dosimetric difference to trigger a new treatment planning if necessary. AS methods are used to contour OAR [37,50,51] and target volumes [52] either differentially or jointly [53,54]. The efficiency of target volume segmentation is a current challenge. For tumor and lymph node areas [55], MR images are preferred (mainly because of image contrast) [56].

Historically, there have been two main types of automatic delineation methods: methods without prior knowledge, and methods with input data. AS methods without prior knowledge can be divided into 3 categories: segmentation by regions [57], by contour detection, and by intensity threshold. AS methods with input data are atlas based auto segmentation (ABAS) [54], statistic model methods, deep learning (DL) methods [58], or hybrid methods [59]. ABAS methods use prior knowledge that is provided by one or more atlas(es) typically using DIR to perform automatic delineation. The statistical model method uses shape variations and appearances of the structures of interest to train the statistical model to perform self-segmentation. DL methods extract features from an input database (images) using neural networks. A step of training (with reference delineated volumes) is needed for DL methods. DL algorithms are based on neural networks consisting of several layers [60]. Hybrid methods can merge one or more algorithms to eliminate their weaknesses and improve the accuracy of the segmentation while having an optimised delineation time. Hybrid methods are frequently used because of their adaptability and flexibility [61].

Lim et al. compiled studies about ABAS algorithms performance from 2005 to 2015, in terms of contouring time and accuracy [52,53]. The majority of the identified work uses manual contouring as a benchmark, in comparison with segmented contours by ABAS algorithms. ABAS algorithms significantly reduced contouring time by 30–60% compared to manual segmentation in a majority of cases [53,54,62]. Several publications evaluated the AS quality with DL methods [63,64]. With parotid gland segmentation they obtained an average DICE index value of 0.85 with a fully convolutional neural network (FCN) [65], and 0.88 with a convolutional network (CNN) [64]. Generated contour from any AS tool need to be validated (edited if necessary) by a human.

Van Rooij et al. [66] provided a new insight into the evaluation of segmentation methods. They compared DL methods to manual contouring. They carried out a geometric evaluation combined with an evaluation of the dosimetric impact with these two types of contours. Despite the variations observed with the DICE index, the effect on the final delivered dose was limited. Despite the need to systematically validate these contours, this raises the issue of performing minor correction operations on a quality segmented contour.

AS could be particularly useful for ART strategies. Indeed, during RT large anatomical variation may occur during the treatment. For target volumes, a GTV shrinkage is quantified as between 17 % and 93 % at mid-treatment (1–3, 67–72) and nearly half of the volume (from 21 % to 75 %) at the end of the treatment [73–75]. To avoid unexpected recurrence, it is necessary to adapt the GTV and the CTV based on anatomical barriers, and not on a geometrical approach.

AS methods allow contouring target volumes such as GTV or Biological Target Volume (BTV) from the information contained in functional imaging (MRI or PET). Foster et al. [76] highlighted the technical difficulties of evaluation between different AS methods using PET images. Image segmentation specific features from functional information need standardisation to obtain robust data for acquisition and reconstruction particularly in the context of multi-centric studies.

The quality assurance of a clinical trial includes a benchmark to assess the contouring quality and is essential to evaluate accuracy and robustness of methods used. Performance levels should be required using indexes and their associated tolerances [20]. Mattiucci et al. consider, as an example, that a mean DICE index value of 0.8 is
acceptable to qualify a benchmark [77]. We recommend to define the metric tolerance threshold used for contour comparison from literature, data according to the volume (OAR, CTV, etc.) evaluated and inter- and intra-operator variability when available. It is also necessary to rely on the geometric parameters of the image (matrix size, slice thickness, etc.) for distance metrics. We therefore recommend to use complementary metrics for pre-clinical evaluation of segmentation such as the DICE index associated to a distance metric (Hausdorff distance maximum or mean distance conformity) [78]. We do not recommend using only the DICE metric, it has limitations due to its definition (overlap between two structures) and it is not adapted for small structures. In the current context where it is still difficult to obtain an acceptable common quality, Cardenas et al. proposed to use a second independent AS software as a secondary check [49].

**Practical recommendations:**

- For AS initial commissioning, a geometric and dosimetric evaluation of segmented contours in comparison with reference contours is recommended.
- For AS evaluation, a combination of several indexes, at least overlapping and distance evaluation metrics with complementary specificities, is required.
- Acceptability threshold of metrics must be defined in function of segmented volume (taking into account inter and intra operator variability, geometric image parameter, etc.).
- Benchmarking AS solution is required to assess contouring quality to know the algorithm performance.
- A systematic verification and human validation by a skilled operator of the segmented contours is necessary.
- For OARs and target volumes, standardisation of structure names and rigorous descriptions are required.
- To avoid unexpected recurrence, it is necessary to adapt the GTV and the CTV based on anatomical barriers, and not on a geometrical approach.

4. Estimation of the delivered dose of the day

4.1. Dose calculation with a TPS

3D imaging (kV-CT, kV-CBCT, MV-CT, or MRI) acquired for repositioning can be used for dose calculation purposes. Thus, recalculating the initial plan on a daily 3D image makes it possible to quantify the dosimetric impact of significant changes in patient anatomy during the treatment course.

Dose calculation from in-room kV-CT can be performed in the same way as dose calculation based on planning kV-CT.

Kv-CBCT-based dose calculation is a complex issue compared to kV-CT-dose calculation (the current reference practice) because of three main features: the imaging quality, the limited acquisition field-of-view (FOV) and theHU consistency. Currently, kV-CBCT images are still prone to artefacts, caused by the lateral scatter, although some of them can be improved by more advanced scatter correction such as iterative reconstruction algorithms [79]. Moreover, limited FOV sizes can lead to an incomplete acquisition of body contours which may be problematic for target volumes (nodes) or shoulders shape required for dose calculation.

The major issue in kV-CBCT-based dose calculation is the HU consistency. Indeed, kV-CBCT HU can vary according to several parameters [80]. Several methods to perform kV-CBCT dose calculation have been proposed in the literature [80,81]: (a) calibration curve between HU and electron or mass densities (HU-D curve), (b) density assignment method (DAM) (c) DIR between kV-CT and kV-CBCT and (d) machine learning to generate a pseudo-CT (pCT).

(a) The HU-density curve established from a kV-CBCT image can be used to convert CBCT HUs to densities for dose calculation. This curve can be defined with either an “adapted” phantom (according to anatomical localisation) or patient kV-CBCT images. Although these methods are straightforward, they are sensitive to kV-CBCT artefacts and patient scattering. (b) The density assignment method (also known as the bulk density method) involves segmenting an image into two to six tissue classes (e.g., soft tissues, air and bones) before assigning density to each class. Nevertheless, this method depends on structure segmentation and provides an image with homogeneous tissues. (c) By deforming kV-CT to kV-CBCT, a “deformed” CT is generated and can be used for dose calculation. CT-CBCT DIR can be difficult due to intrinsic kV-CBCT limitations, such as noise, low contrast, and reduced FOV. However, in ART workflow, patient position is maintained and anatomical changes are typically small, which is favorable for DIR process. (d) Machine learning methods are based on patches or DL methods (DLMs), to generate a pCT (i.e. synthetic images) from kV-CBCT. Machine learning methods require a large training cohort and most of these methods require coregistered data for the training step. Some studies proposed to compare several methods for H&N CBCT-based dose calculation. Glacometti et al. compare HU-D curve, DAM and DIR method for H&N VMAT plans [82]. In this study, the DIR method provides the best agreement considering the relative dose difference between kV-CBCT and kV-CT doses delivered to 99% (D99%), 95% (D95%) and 1% (D1%) of the tumor and nodal PTV. Maximal dose differences of 3.9%, 3.2% and 2.7% for HU-D curve, DAM and DIR method respectively, were obtained for D99% of nodal PTV [82]. Barateau et al. compare the three previous methods with a deep learning method (DLM) for H&N RT [83]. The mean 3D gamma pass rate (local, 2%/2 mm, 30% dose threshold) was 91.0 ± 5.3%, 97.9 ± 1.6%, 98.8 ± 0.7% and 98.1 ± 1.2% for HU-D curve, DAM, DIR and DLM respectively. They conclude that for H&N RT, DIR and DLM appear to be the most attractive methods in terms of dose accuracy as well as calculation time [83].

The MV-CT dose calculation can be performed with a specific curve calibration. Compared to kV-CBCT, MV-CT has the advantage of being independent of the patient’s corpulence thanks to the collimation. Stability of calibration curve over time depends on linac target [84,85]. Zhu et al. showed the impact of the acquisition and reconstruction parameters on dose calculation [86]. Monthly monitoring of the calibration is recommended as well as after any modification of the target or beam [87–89]. The dose can be recalculated with an accuracy of about ±2.5% for H&N with a HU-ED curve [90]. Other methods such as DIR (between kV-CT and MV-CT) [91] or DLM have been proposed [92]. MV-CT images have some artefacts, called zipper artefacts, due to the isocenter misalignment, but their dosimetric impact was found insignificant [93]. Moreover, due to high energy, metal artefacts are considerably reduced compared to kV-CT or kV-CBCT. This represents a major advantage in case of dental or orthopaedic implants for patient positioning and dose of the day calculation [94]. MV-CT should be preferred in such cases.

The main limitations of MV-CT imaging are the poor contrast due to high energy and the restricted FOV size. With such imaging modality, parotid glands are not visible [95]. These images can be used to assess the dosimetric impact of significant weight loss or target volume reduction. A study evaluated the impact of weight loss and anatomical change during H&N RT, measuring differences between planned and delivered dose to spinal cord based on MV-CT images [96]. They obtain for 133 H&N patients a slight impact on the dose delivered to the spinal cord, with an absolute dose difference of 0.9 Gy (95% CI 0.76 to 1.04 Gy) [96]. This study shows the feasibility to use daily MV-CT images for dose calculation and indicates that the spinal cord is not the critical organ to consider for H&N ART approach. Other approaches could be used instead of kV-CT for replanning. MRI replanning can be performed with an independent MRI-only workflow or using MRI-linac. The main issue of MRI (re)-planning is the lack of electron or mass density information, necessary for dose calculation. To overcome this issue, various methods were developed to generate Synthetic Computed Tomography images from MRI [97]. For MRI-based dose calculation, recent DLMs provided dose uncertainties lower than 2% compared with the dose calculated on the corresponding kV-CT [98–100].

Some commercial softwares allow calculation of “dose of the day”
5. Dose monitoring (dose accumulation)

Dose monitoring can be used to trigger a replanning, to replan taking into account the dose previously delivered, or to estimate the delivered dose. Dose monitoring can be performed with dose accumulation of each “dose of the day”. Commercial solutions (integrated in TPS or not) are now available to perform dose accumulation during or after the treatment.

Dose distributions need to be in the same space to be added. In a first approximation RIRs between images of the day and planning image can be performed, and the dose distributions corresponding to each image can be added. This method does not take into account the deformation, that is why dose warping is needed before the dose accumulation. However, deformed dose is not measurable and dose received by cells in case of tissue appearance or disappearance is still an unresolved problem.

The propagation of large deformation errors leads to large dosimetric errors. The result of the deformation and dose accumulation is subjected to several influence parameters: image quality, parameters related to DIR methods, amplitude and type of deformation, dose matrix (voxel size, dose gradient) [114]. Vega et al. evaluated several DIR algorithms from a daily dose calculation perspective on kV-CBCT images: 9% (SD = 4%) of the voxels differed by more than 2% from the prescribed dose [115]. In high gradient regions this value increased to 21% (SD = 6%) and in regions with poor image quality to 28% (SD = 9%). Qin et al. highlighted the need for biomechanical module for large deformations and high gradients [116]. Rigaud et al. evaluated several DIR methods and found that uncertainty on cumulative parotid mean dose was 4 Gy on average (SD = 2.27 Gy) for a prescription dose of 70 Gy (2 Gy per fraction) [117]. Pukala et al. evaluated five commercial products and pointed out that the effects of registration errors on a dose volume histogram (DVH) are not easily predicted. In fact, the amplitude and direction of structure registration errors, dose gradient and distance from structures influence the quality of the dose deformation on the structures [118]. Glide-Hurst et al. recommended a dosimetric accuracy of ±5% on the entire ART workflow (DIR, segmentation, dose calculation, and dose accumulation) [119,120]. They emphasized being cautious with dose warping and dose accumulation when applying ART to decision-making.

Practical recommendations:

- Evaluation of RIR and DIR performance (part 1) is recommended in order to know algorithm behaviour in homogeneous and heterogeneous medium, and provide an order of magnitude for dose uncertainties.
- Be aware of uncertainties due to registrations and dose warping.
- It is recommended to perform localised registrations for dose accumulation in case of specific area of interest (dose to spinal cord for example).
- To trigger replanning, approximations can be done by accumulating weekly doses instead of daily doses.

6. ART workflows

Several ART strategies were described in publications as offline, online, hybrid or in real-time [120]. In H&N RT treatment, offline and online ART can be employed. Whereas replanning process in online ART implies the use of in-room 3D-image, the process in offline ART is based on new image acquisition (the standard being a new CT planning) during treatment. The delineation of tumor(s) and OAR(s) is needed on this image (Part 2). As the prescribed dose was partially delivered, this dose contribution can be taken into account and is called “background dose”. With this background dose, an estimation of the total delivered dose can be performed. The background dose must be cautiously used for ART since it is based on dose warping which has uncertainties and limitations (part 4). As anatomical changes are a gradual process during H&N RT treatment, at least weekly 3D in-room imaging is recommended to determine if an ART procedure is necessary [121,122].

6.1. Offline ART workflow

Fig. 1 presents a typical offline H&N ART workflow.

The first step is to define a criteria that can be used to trigger a replanning. They can be visual based daily image analysis (weight loss, parotid glands position and volume decrease or tumor shrinkage, positioning issue, etc.) or quantitative (PET, estimation of the daily delivered dose, ...). There is no consensus on these criteria [5,118]. A recent review gathers studies on this topic [123]. Decision criteria is a real issue and could be simple or complex: Richter et al. include patients in a highly automated workflow with the observation of anatomical changes...
on kV-CBCT during treatment as a decision criterion [124]. Li et al. propose a complex method using functional imaging and kV-CBCT, combined with radiomic and genomic analyses, to identify the directions of escalation or de-intensification of treatment [125]. Castelli et al. have established a nomogram to identify, from the first week, patients likely to have a significant dose increase in the parotid glands [126]. Grepl et al. evaluated the dosimetric benefit to adapt dosimetry based on weekly MR images to reduce dysphagia for patients treated by chemo-RT [127]. The legitimate fear of over-irradiation of the spinal cord tends to be a criterion for the implementation of ART; however, Noble et al. analysed that anatomical changes during treatment do not lead to a risk of spinal cord complications [96]. In vivo portal dosimetry can help to detect anatomical and/or dosimetric modifications during treatment course [128]. All these elements of analysis have different relevance and effectiveness. Brouwer et al. explored the literature to identify selection criteria and concluded that there was a lack of data quality and the need for further investigation to establish relationships with clinical outcomes [5]. Thus, there is currently no solid data on the criteria for choosing a re-planning.

Available data show a strong impact of early replanning (1st or 2nd week of treatment) [129,130]. In H&N cancer, anatomical variations appear progressively over the first week of treatment and continue throughout the full treatment. Early replanning allows to correct these variations for a higher number of fractions compared to a later replanning (4 or 5th weeks of treatment). Hence, due to the progressive nature of these anatomical variations, off-line ART seems useful, limiting the risk of variability linked to daily replanning and this despite the lack of current automation of replanning tools at the moment.

Furthermore, different studies have shown that the use of robust optimization (RO) instead of conventional treatment planning could be a way to avoid or minimise re-planning during the treatment course while better sparing the OARs. The principle of RO has been widely described elsewhere [131]. In RO, the optimization constraints are applied on the CTV, while the treatment uncertainties are directly incorporated in the objective function of the optimization algorithm. In H&N cancer, several studies have shown that RO led to improved target coverage and reduced dose to the OARs compared to conventional VMAT using PTV strategy. Thus Wagenaar et al. showed on a cohort of 10 patients with various H&N tumor localisations treated in VMAT, that RO resulted in a significantly more important D98% to the CTV but significantly lower dose to the main OARs than PTV-based planning [132]. Miura et al. performed similar work on patients with larynx cancer treated and reached the same conclusions. They also observed that to perturb the nominal dose distributions conducted to less important dose variations to the CTV and OARs using RO [133]. In an original way, Cubillos-Mesias et al. found that to include the first 2 weekly kV-CBCTs in the RO process of intensity modulated proton therapy planning enabled to better take into account anatomical changes than conventional RO or PTV-based strategy [134]. They thus showed that PTV-based and RO approaches were not sufficient to take into account anatomical changes in 10 and 5 out of 20 patients, respectively, resulting in the need for plan adaptation. Conversely, using anatomical RO, in all except one patient the CTV coverage was conserved and no adaptation was required. Hence, RO has the potential to provide treatment plans that are more robust to anatomical changes than PTV-based planning. Finally, both planning strategies showed improved plan robustness when considering the anatomy of the two first weeks of treatment.

6.2. Online ART workflow

Currently, in most cases, the adaptation during the H&N RT course is carried out offline. However, adaptation software are now commercially available to perform an “online” adaptive replanning [128,135]. This software benefits from specific developments with automatic or semi-automatic solutions. Commercial online solutions (kV-CBCT-based [136], MR-based [137]) offer automated tools for assessment, and replanning (if necessary or pre-determined) to have a fraction as short as possible for the patient (on the treatment couch during the plan adaptation). These processes can be interesting not necessarily for a
systematic daily replanning, but when needed (or weekly for example) during the treatment course when one (or more) replanning decision criteria are met. This workflow is still being evaluated and is the subject of clinical trials (NCT05666193, NCT03972072 (MARTHA-Trial)). Systematic daily ART may not be required in H&N cancer. Indeed, the labour cost of this approach seems to be unfavourable. This opinion is to be qualified for moderate or extreme hypofractionated treatments [138], because that potentially reduces the work on the whole treatment. This concept remains to be evaluated.

An online kV-CBCT-based system mounted with a ring linear accelerator is currently available to implement an automated online ART workflow with the patient on the treatment couch. This platform offers the possibility to treat in a standard way (with daily kV-CBCT) and also to access to dose accumulation on initial kV-CT after every session for follow-up. Another opportunity to treat with this kV-CBCT-based system is to adapt online dosimetry to new anatomy in an integrated way. A timing and automation study is carried out by Yoon et al. [139] with online ART workflow with an “in silico” study, median time session is around 19 min and provides preliminary data on the possibility of carrying out these treatments.

Recent MRI guided linear accelerator (MR-linac) devices allow MR acquisition before each treatment delivery. These images can be used to perform dose monitoring or replanning in a context of MR-guided H&N ART [140,141]. For now, H&N localization is not the main tumor localization for MR-guided RT but some clinical trials are ongoing [45]. As presented in Fig. 2, workflows should be adapted for MR linac online H&N ART.

Chen et al. reported their initial clinical experience with a 0.35 T MR-linac machine for 18 H&N patients and proved that MR-guided RT can achieve clinical outcomes comparable to those traditional IMRT for H&N cancer [142]. Lim et al. compared planned and delivered doses after an adaptation to position workflow (“adapt to position” = shift of isocenter and MLC based on daily MR scan) for 8 H&N patients with a 1.5 T Elekta MR-linac device [143]. They found that the delivered dose to some OARs was significantly higher than original planned dose, suggesting that an adaptation to shape process (“adapt to shape” = full plan reoptimization based on re-contoured daily MR scan) should be used in some specific H&N cases.

Mc Donald et al. described their institution workflow for a H&N treatment with 1.5 T MR-linac and demonstrated the treatment feasibility for 10H&N cases [10]. A T2-weighted MRI was acquired for daily setup verification and also used for plan adaptation. Registration was done between MRI of the day and reference image. They choose a 5 mm isocenter shift as criteria to choose between “adapt to position” (shift < 5 mm) or “adapt to shape” (shift ≥ 5 mm) process. In this study, for efficiency reasons, “adapt to position” workflow was performed offline (to create a new reference plan). In case of “adapt to position” online workflow, MLC position was adapted to be consistent with reference plan in terms of dosimetry criterions (PTV coverage and OAR constraints).

6.3. Workflow organization

A specific ART H&N workflow should be defined and validated in each institution, including method, imaging frequency, decision criteria and professionals concerned (RTTs, physicians, medical physicist, etc.). The number of replanning and selection criteria must therefore be clearly defined before any ART approach is taken. Participant roles should be well defined for all long treatment steps as all treatment phases (delineation, treatment planning, validation, quality control) should be realised in a reasonable time. Some steps could be automated (delineation, adaptation of treatment plan, etc.) permitting a new planning in a shorter period. Expertise in the assessment tools and those used for ART is essential. Regardless of the strategy required, automation is recommended in particular for recontouring and dosimetry, with systematic human validation and corrections (when necessary) for delineation. Tools such as DIR, automatic image segmentation, and expert planning software can help to reduce both operator variability and human time requirement and can significantly reduce ART process.

In particular, the use of auto-planning solutions could help to achieve

![Fig. 2. MRI online H&N adaptive radiotherapy (ART) workflow.](image-url)
these objectives, the ultimate goal being efficient generation of high-quality plans. Among the solutions are the automated rule implementation and reasoning (ARIR), a posteriori and a priori multicriteria optimization (MCO) and knowledge-based planning (KBP), which can be divided in two additional categories: statistical modeling of case/atlas-based, and machine learning methods (DL) [144]. For H&N cancer planning, atlas-based KPB and ARIR approaches were shown to provide clinically acceptable plans, which were comparable with those obtained manually [145,146]. Despite difficulties linked to the large amount of homogeneous data required for the database construction, DL-KBP solution has the potential to drastically reduce the planning time but still needs further investigation to be more efficient in providing plans of similar quality to manual plans [144]. A posteriori MCO algorithm which proposes to choose one solution among several displayed on a “Pareto surface”, was also efficient in providing clinically acceptable plans, but higher effective working time and optimization time were reported compared to KPB and ARIR [147]. In contrast, a priori MCO directly and automatically generates a single Pareto-optimal plan using a wishlist with predefined dosimetric goals. Overall superiority of this algorithm compared to manual plans was recently demonstrated for H&N cancer but with an optimization time (1 h) which was not compatible with an online ART workflow [148]. Recently, Archambault et al. introduced the Intelligent Optimization Engine (IOE), a semi-automated TPS specifically designed for online ART [136]. This solution is similar in some points to a priori MCO used to obtain the initial planning and subsequently used for the automated plan generation on couch. Several studies showed the ability of IOE to generate plans with a quality comparable to manual planning within a short time-frame (<10 min) for pelvic [149,150] and H&N cancers [139].

If specific resources are required for offline ART strategies, they will most often use “conventional” care processes but in a shorter time frame than that used for initial preparation.

With online workflow, the institution organisation needs to be adapted compared to a standard RT workflow particularly in terms of staff, training, and assessment. This also implies thinking about RTT therapists and dosimetrists roles according to the organisation [151]. Online ART requires specific medical and paramedical resources at the treatment machine in a choreography organised and prepared in advance with, in particular, prior training and validation of the skills of each operator through specific accreditation. The presence of the physician for delineation review may be necessary but it can be delegated to the RTTs mainly for OARs [152] and in certain situations and after training with certification for target volumes [151,153]. In all cases, this implies specific training and skills, with a credentialing for RTTs particularly. However, treatment responsibility will always rely on physicians and physicists. To delegate responsibility, training courses have to be formalised and competences should be validated by the responsible (physician for contouring and physicist for planning).

**Practical recommendations:**

- Prior training and skills validation are required for each operator through specific accreditation, for offline and online ART treatments.
- To delegate, training courses have to be formalised and competences should be validated by the responsible (physician for contouring and physicist for planning).
- As anatomical changes are a gradual process during H&N radiotherapy treatment, at least weekly 3D in-room imaging should be recommended to determine if an ART procedure is necessary.
- Online strategies can be interesting for H&N ART, not necessary for a systematic daily replanning because progressive variations are observed but needed during the treatment course when one (or more) replanning decision criteria are met.
- Robust optimization should be used to improve plan robustness to anatomical changes.

### 7. Quality assurance

#### 7.1. Machine and imaging quality assurance

As in non-ART workflows, a comprehensive machine QA program is needed. QA for conventional and non-conventional linacs has been extensively described in the literature and one should refer to national or international (specific task group reports) guidelines [154–156].

As imaging is an important tool for H&N ART, image quality needs to be particularly assessed for any clinical use. A balance between image quality and imaging dose is also required. As CT scans, in-room imaging protocols need to be optimised. In France, a periodic quality control is mandatory for the CT-scanner and recommended for kV-CBCT [157] and MV-CT in-room imaging devices. Spatial resolution and contrast are key components of these controls. Some quality controls need to be performed according to the image usage. As an example, stability of HU of kV-CBCT, MV-CT, or kV-CBCT have to be regularly checked if dose calculation is based on such images. In addition, coincidence between imaging and treatment isocenters and accuracy of treatment table displacements must be checked periodically to make sure that treatment could be performed consistently with image registration.

For MRI, homogeneity and distortion in the magnetic field needs to be periodically checked as well as some geometric parameters [158]. For PET imaging, thresholding and reconstruction protocols need to be assessed [159].

Except for MRI, in-room imaging involves an additional irradiation. Imaging dose has to be quantified and the protocols have to be optimised (according to clinical purpose and patient morphology) to reduce the imaging dose while ensuring that the ART strategy objectives remain consistent. Indeed, imaging dose can be reduced up to a factor 10 with optimised protocols [160,161].

#### 7.2. Patient specific quality assurance (PSQA)

Since its implementation in the mid 1990’s, guidelines for commissioning and QA of IMRT have been widely described and documented [162–164], and performing Patient Specific Quality Assurance has been recommended by several professional organisations [163]. Miften et al. provided specific recommendations to implement measurement based-PSQA and methodologies to establish tolerance and action limits [165]. They particularly recommended: to perform IMRT QA measurements with a True Composite method (ie. using a stationary QA device placed on couch and measuring the actual plan of the patient), to analyse the dose comparisons in absolute dose mode, to perform a dose calibration of the QA device before each measurement session and to use a global normalisation for the $\gamma$ index analysis.

While pre-treatment and offline measurement based-PSQA can be done practically in offline ART, it cannot be implemented with a patient on couch in an online ART workflow. In this case, the PSQA will rely on a surrogate system, typically independent computer calculation. This solution will provide a quick result compatible with the timescale of an online ART session. If independent recalculation has been deemed capable of performing a reliable PSQA [166], it has never been formally proved as a sufficient surrogate for measurements. Indeed, independent recalculation will only catch calculation errors such as beam modelling inaccuracies, and not errors in delivery or machine output. Furthermore, in order to confidently use it, it would be desirable to check the correlation between calculation errors and the surrogate results by studying its sensitivity [167].

The risk in online ART is that the adapted plan deviates from normal operating conditions leading to an unacceptable plan for the patient. A good approach would be to monitor the complexity of adapted plans to check that they do not fall outside defined boundaries. Tolerance and action limits should be based on local experience. Indeed, several publications showed that treatment plan complexity does not predict PSQA performance at multi-institutional scale as complexity metrics and their
correlation with PSQA results are highly dependent on the material used (TPS, linac, measurement device, planning technique) [168,169]. Zhao et al. evaluated the need to perform a measurement based-PSQA for adapted plans during online ART. They concluded that measurements may not be necessary for every adapted plan and suggested to perform periodic checks to monitor the trend of PSQA results [170].

7.3. End-to-end testing and risk management

Before each clinical implementation of a new radiotherapy treatment technique, an end-to-end (E2E) test should be conducted. This is particularly true in the case of ART workflows where each specific component has to be checked individually as well as the complete process. Recommendations about QA of each component are addressed in the previous sections.

The aim of E2E QA is to check the accuracy of the delivered dose to the patient under real-world conditions, including ultimately verification of accumulated dose [120,171]. However, the challenge is that, currently, no commercial phantom complies with the requirements of E2E testing of the whole ART workflow. Only a few in-house physical phantoms with embedded organ shape deformation have been designed by independent centres, for H&N [172], pelvic [173,174] and lung [129] regions, demonstrating the feasibility of E2E ART QA.

As already mentioned, using in vivo portal dosimetry or fluence detectors is another interesting tool that should be used to control accuracy between planned and delivered dose. Lim et al. found a statistically negative correlation between variation of transit fluence and variation of volumes on CBCT. Such in vivo tools could help to decide if replanning is necessary [108].

In addition, as implementation of new treatment techniques yields new potential risks, a failure mode and effect analysis (FMEA) is highly recommended as part of a risk management program to target potential pitfalls in the process [175]. Rippe et al. conducted a FMEA of the adapted planning process on a MR-Linac, and found that a large number of error sources, such as missing slices in volumes or erroneous margin extension of target volumes, were not covered by the available QA tools provided by the manufacturer. They hence developed an additional software to fill the gaps [176].

**Practical recommendations:**
For machine and imaging QA:
- A comprehensive machine QA program according to national or international guidelines with a particular focus on imaging should be designed.
- The assessment of image quality and image-guided ART protocol optimisation (particularly on acquisition and reconstruction parameters) is recommended.

For the plan QA:
- A systematic measurement of pre-treatment PSQA should be done following recommendations made by Miften et al. [165] (before the first treatment session in an online ART workflow and for each new plan generated in an offline workflow).
- For online ART:
  - A sensitivity analysis to errors of the surrogate QA system should be conducted to assess its reliability.
  - Complexity of adapted plans should be monitored in order to identify plans deviating from normal operating conditions.
  - Periodic measurement-based PSQA of adapted plans should be performed to identify a potential drift in plan QA.

For the end-to-end testing and risk management:
- Each component of the ART workflow should be assessed with a specific QA.

8. Conclusions

This paper gives an overview of the key elements of H&N ART and practical recommendations. This manuscript and in particular the recommendations are intended to assist and provide the minimum requirements for using these tools and strategies when implementing them in practice or in clinical trials linked to ART.

A particular attention should be drawn on the increasing use of artificial intelligence (AI) in radiotherapy. Indeed, AI-based applications, such as automatic segmentation, synthetic CT generation or automatic planning, allow a significant gain in efficiency in offline and online ART workflows but may appear as black boxes [60,177,178]. Claessens et al. provided practical methodologies for the commissioning, implementation and QA of AI models used in clinical practice [179]. For all H&N RT patients, anatomical changes are observed during treatment, that are generally associated with localisation, patient’s anatomy and response to treatment. Imaging or dosimetric tools offers the possibility to detect variations during treatment, but thresholds need to be defined to detect when these variations mandate a treatment plan adaptation. Automatic tools are needed to catch patients who require plan adaptation. In the current state of our knowledge, daily online adaptation for H&N RT normofractionated treatments is not recommended but rather on an individual basis with regular monitoring of observed changes. ART strategies are an opportunity to optimise treatments when clinically it is an issue. ART strategies could adapt dosimetric parameters keeping target coverage and OAR sparing but could also be used to increase local control using dose (de)escalation. This latest objective can be achieved thanks to anatomical imaging allowing tumor visibility or thanks to PET or MRI functional imaging which allow for increased personalisation of the patient’s response to treatment.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References


[49] Carpentier CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in Auto-
semradonc.2019.02.001.


[51] van der Veen J, Gulyan B, Nuyts S. Interobserver variability in delineation of


[53] Lahoud L, Mohamadi T. A comparative study of Image Region-Based


[56] Giacometti V, Housseln AR, McGarry CK. A review of dose calculation approaches

[57] van Rooy J, De Veire R, Lantin M. A Phantom
Evaluation of the radiobiological impact of anatomic modifications during
radiation therapy for head and neck cancer: can we simply summarize the dose?
radioonc.2010.05.009.

[58] Foster B, Bagur U, Manoros A, Xu Z, MoIura DJ. A review on segmentation of


[60] Jena R, Kirkby NF, Burton KE, Hoole AN, Tan LT, Burnet NG. A novel algorithm

reduction using iterative CRRT reconstruction algorithm for head and neck

jmiro.2019.08.038.

jmiro.2019.08.038.

[64] van Rooy J, De Veire R, Lantin M. A Phantom
Evaluation of the radiobiological impact of anatomic modifications during
radiation therapy for head and neck cancer: can we simply summarize the dose?
radioonc.2010.05.009.

[65] Lalaoui L, Mohamadi T, A comparative study of Image Region-Based


[67] Barateau A, Garfopecas C, Cugny A, De Figureiredo BH, Dupin C, Caron J, et al. Dose calculation accuracy of different image data to density tables for cone-


tomographic scan methodology on setup verification and adaptive dose


