Clinical Oncology

Guidance on auto-contouring in radiotherapy

NOVEMBER 2024

The Royal College of Radiologists

Contents

This guidance is endorsed by the Institute of Physics and Engineering in Medicine's Radiotherapy Special Interest Group (RT-SIG)

Executive summary

There is rapidly increasing availability of auto-contouring systems for use in radiotherapy planning and treatment. These have the potential to improve treatment quality and consistency, and to reduce the treatment pathway time. This guidance has been developed to support clinicians with the use of auto-contouring in clinical practice.[1]

There are several stages in the pathway of auto-contouring implementation in which clinical oncologists may have differing levels of involvement. This document addresses the important aspects of auto-contouring commissioning, implementation and ongoing clinical use. The most important principle throughout auto-contouring applications is that the healthcare professional approving any auto-contours is ultimately responsible for their clinical use. Clinicians need appropriate training since reviewing auto-contours is a different skill to manual delineation. The reviewer should use a structured approach for checking contours and prioritising areas of critical dosimetric importance, and have dedicated time in their job plan for review and editing.

Commissioning by a multidisciplinary team should ensure the safety and effectiveness of the auto-contouring system using local data, representative of the real-world clinical cohort. Ongoing surveillance of the performance and clinical utility of the auto-contouring system after clinical implementation must be undertaken. Clinician training will need to stay up to date as this is a rapidly evolving field and the scope of auto-contouring is likely to change.

Terminology in this guidance

This guidance has been developed to aid all healthcare professionals involved in the use of auto-contouring systems for radiotherapy treatment. While the clinical oncologist, or radiation oncologist, usually has overall responsibility for a radiotherapy treatment, there can be delegation of responsibility with advanced practice roles carried out by medical physicists, dosimetrists and therapeutic radiographers. This is likely to expand with increasing use of technologies based on artificial intelligence. Therefore, when the term **clinician** is used in this guidance, this will include all healthcare professionals appropriately trained for the task.

There are various terms used in the literature related to auto-contouring in radiotherapy. For consistency throughout this document, standardised wording is used but it is acknowledged there may be alternative or preferred wordings for the same or similar phrase. **Auto-contouring** is the automatic production of contours on an imaging data set, which can be achieved using artificial intelligence approaches (eg deep learning) or non-artificial intelligence approaches (eg atlas-based). Alternative terms include **auto-segmentation** and **artificial intelligence-based auto-segmentation (AIBAS)**.

The technology that automatically delineates structures is an auto-contouring system, and alternative terms include **auto-contouring model**, **auto-contouring tool**, **auto-contouring solution** and **auto-contouring pathway**.

01 Introduction to auto-contouring

- **The healthcare professional approving auto-contours is ultimately responsible for their clinical use.**
- **Auto-contouring performance is influenced by how a system is trained and by how it works. Auto-contouring systems should only be applied in the context of their intended use.**
- **Clinicians should understand that the performance of auto-contouring systems may vary for different tumour types and across different imaging modalities. The impact of using auto-contours for target structures compared with organs at risk (OARs) will differ.**
- **Clinician training will need to stay up to date as this is a rapidly evolving field and the scope of auto-contouring is likely to change.**

Auto-contouring in radiotherapy is a rapidly evolving area of innovation. In September 2023, the National Institute for Health and Care Excellence (NICE) approved nine radiotherapy auto-contouring providers with the stipulation that contours must be reviewed by a healthcare professional prior to clinical application.[2] Any healthcare professional approving auto-contoured structures is ultimately responsible for their use. This is usually a clinical oncologist; it may also be a delegated professional such as a therapeutic radiographer or dosimetrist based on local IR(ME)R protocols.[3] Clinicians must therefore have appropriate knowledge and training on using and assessing auto-contours.

Auto-contouring refers to a technology that automatically delineates structures for radiotherapy planning. This may include both target volumes and OARs and is becoming available for an increasing number of tumour types. It has the potential to improve the consistency and quality of treatment and to reduce the treatment pathway time, although at present evidence is limited.[4,5] Auto-contouring systems are medical devices that are subject to strict regulatory processes, and departments should familiarise themselves with the relevant medical device regulation guidance.[6,7,8,9]

There is demand from clinicians for further guidance on how to use this technology.[1,10] Although healthcare professionals do not need to know the intricacies of how to create an auto-contouring system, they should understand the processes involved in implementing and using a system safely on a day-to-day basis. Operator error has been demonstrated to be one of the biggest causes of error when using auto-contouring in the clinic.[11]

01 Introduction

Established methods for producing auto-contours include atlas-based and deep learning.[12] Atlas-based systems use 20 or more previously delineated cases to build a set of atlases. An auto-contour is produced by applying the contours from the atlas onto a scan using deformable image registration. Deep learning methods produce original contours based on patterns the system has learned during model training. Large numbers of pre-labelled scans are required to make '**training data**'. Each voxel on a training scan is allocated to a region of interest and the scans are then entered into a deep learning system made up of neural networks.[12] Neural networks are able to 'learn' patterns and predict how voxels will be labelled on a new scan. Applying a trained system to an unlabelled scan will then result in an original set of delineations, based on the patterns that have been learned.

The quality and quantity of data used to train auto-contouring systems directly affect performance.[13] Any auto-contouring system for clinical use should have been trained on data similar to the intended clinical purpose. The training data should have been contoured with the same protocol and on the same imaging modality and patient position. Similar patient demographics are also required, including but not limited to age (adult versus child), gender, body habitus, and presence or absence of previous surgery or metallic artifacts. There are increasing numbers of commercially available and in-house auto-contouring solutions. How the system has been trained is an essential consideration when choosing between options, and departments should consider asking commercial vendors to disclose this information.

All auto-contouring systems need to be appropriately commissioned before clinical use to confirm they are applicable to the local patient population.[4,14] Assessments are also required as part of day-to-day quality assurance (QA) and as part of formal post-implementation monitoring. A key aim of post-implementation monitoring is to monitor for a drift in performance and to identify if any **automation bias** is occurring. Automation bias is an important source of error in auto-contouring for radiotherapy, and all clinicians using auto-contouring systems should be aware of it. Automation bias is a phenomenon whereby reviewers eventually favour output from the automated system, despite having evidence or knowledge that would suggest the automated system is wrong.[11]

Another important form of bias is **anchoring bias**, which refers to the way a human observer is unduly influenced by the first interpretation of a data-processing procedure that they see. In this case, the user would observe auto-contoured structures on the image data of a new patient and be unduly influenced as to their accuracy. If the same user performed a manual segmentation of the images first, and then reviewed auto-contoured structures, the accuracy and perceived utility might vary.

This guidance will discuss the different stages of selection, implementation and ongoing monitoring of an auto-contouring system to support the safe clinical use of this new technology.

02 Assessment metrics

- **There are multiple methods to evaluate auto-contouring systems, and the clinical utility of each method can vary. The most appropriate assessment metrics should be chosen for each stage of implementation and clinical use.**

Auto-contouring systems need to be assessed during the different stages of implementation and use to ensure their validity and ongoing safety. These assessments may be performed as part of commissioning, in day-to-day use and for formal post-implementation monitoring. Assessments may also be required by a department choosing which system to implement, or to demonstrate an overall efficiency benefit when producing a business case.

Assessment methods include geometric, time-based, dosimetric and qualitative metrics, which have different strengths and weaknesses as summarised in Table 1.[15]

Table 1. Metrics used for auto-contour assessment

Type of metric	Examples	Strengths	Weaknesses
Geometric	DICE similarity coefficient Hausdorff distance Volume comparisons Surface DICE coefficient Added path length	Easy to understand and calculate	Limited clinical relevance Reliant on 'ground truth' Influenced by structure size
Time-based	Comparison of manual contouring time with auto-contour generation and editing time	Defines utility for a department	May be artificial and simulated Labour-intensive to perform
Dosimetric	Comparison of dosimetry for plans made with auto-contours, manual contours and edited auto-contours	Clinically relevant	Time-consuming Reliant on quality of manual contours
Qualitative (human-centric)	Rating using a Likert scale Blinded Turing test	Reflects usual clinical practice Can be used for bias assessments (eg deliberate introduction of errors)	Vulnerable to bias and operator interpretation Difficult to standardise

Geometric tests are numerical tests that can often be performed within the treatment planning system. They compare an auto-contour with a 'ground-truth' contour and assess characteristics such as overlap, distance between surfaces, volumes and structure centre positions. Newer geometric tests including added path length and surface DICE coefficients may also estimate the amount of manual editing required for a structure.[16,17] Although geometric tests are easy to perform and understand, they have very limited clinical relevance. No single geometric test is accepted as reliable for detecting errors in all scenarios.

02
Assessment metrics

Time-based assessments compare how long it takes to contour structures manually with the time taken to produce, check and edit auto-contours. These studies can be useful to a department if they demonstrate total time savings to support a business case for purchasing a system. However, they are often artificial and simulated, and challenging to complete.

Dosimetric assessments will ultimately determine if the use of auto-contours results in a safe and effective radiotherapy plan. Radiotherapy plans are generated using manual contours, unedited contours and edited auto-contours, with comparison of the dose-volume histograms by transposing each set of contours onto each plan. The clinical significance of contouring errors is difficult to determine from geometric measures alone, so dosimetric evaluation is an important part of auto-contouring evaluation and commissioning.

Qualitative assessment replicates real-life clinical practice when a clinician assesses an auto-contour. Clinician review is the only method that will be able to pick up one-off failure cases of an auto-contouring system and it is crucial that this takes place in the day-to-day use of auto-contours. There are multiple ways to formally perform a qualitative assessment. Likert scales of clinical acceptability or clinician satisfaction are most commonly used. These scales indicate the degree of editing a contour would need prior to clinical use (eg ranging from recontour from scratch to no editing needed). In a blinded Turing test, clinicians are shown a sample of manual contours and automatic contours for the same cases. They are asked to identify if the contour has been generated by a human or computer and which contour is better. If the auto-contours are comparable with manual contours, they will be indistinguishable from manual contours and the correct source will only be identified 50% of the time.[18]

Qualitative assessment is vulnerable to variations in operator interpretation and automation bias. Using standardised assessments and processes can reduce this risk. Performing regular formal post-implementation monitoring can also identify whether performance drift and acceptance of automation bias is occurring.

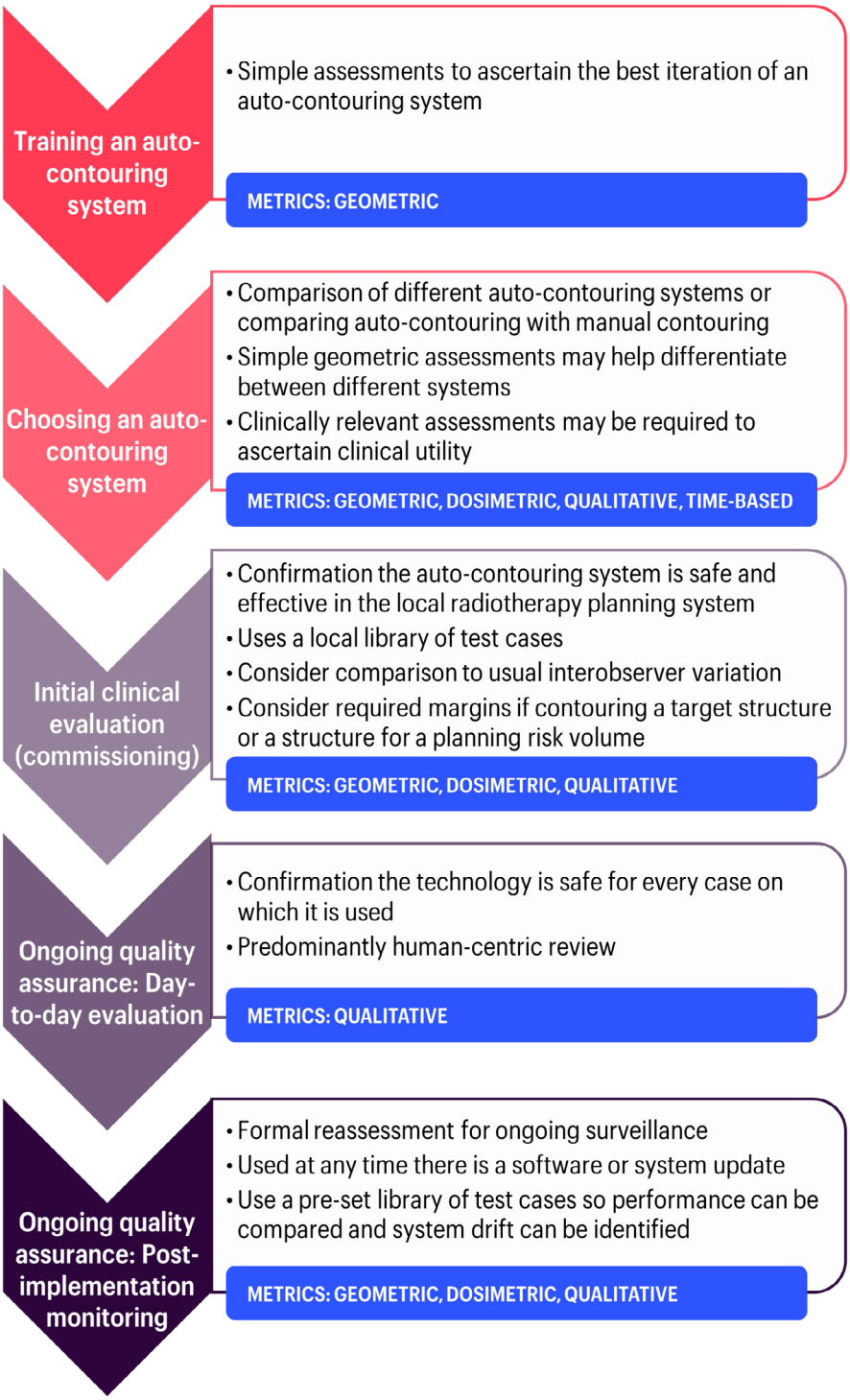
Which metrics to use?

There is currently a lack of consensus on the best methods to assess auto-contouring systems, especially for clinical utility.[18] The assessment metrics used in research are currently very varied making it difficult to produce standardised recommendations. The most appropriate metric to use will depend on the aim of the assessment. Different evaluations may therefore be needed for research, clinical commissioning, QA or business cases. When considering clinical use, clinician involvement is essential and standardised approaches are needed.[19] The different time points when an auto-contouring system may need to be evaluated and the suggested metrics are set out in Figure 1.

A fundamental challenge for assessing auto-contours is that manual contours are only ever a surrogate for ground truth and that delineation in radiotherapy is impacted by uncertainties.[20] Comparing one auto-contour with one ground-truth contour does not take these uncertainties into consideration. Pre-existing guidance recommends assessing auto-contours in the context of usual inter- and intra-observer variation.[14] Considering the range of this variation may be a useful approach to ascertain if auto-contours perform consistently with local clinical practice. The use of standardised contouring protocols and peer review is important when deciding what reference contour(s) should be used for evaluation.

02
Assessment metrics

Figure 1. The recommended assessments for the different stages of auto-contouring implementation and ongoing use



03 Selection of an auto-contouring system

- **The proposed use for an auto-contouring system should be clearly defined to establish the required system specifications.**
- **Available commercial, open-source and in-house systems should be carefully evaluated compared with current manual pathways. Different solutions may work better for specific tumour types, radiotherapy techniques and centres.**
- **When considering an in-house or open-source system, it is necessary to comply with all relevant medical device regulations.**
- **All systems must be subjected to the same rigorous commissioning and QA processes.**
- **The scope of auto-contouring is likely to change in the future and this should therefore be considered when selecting a system.**

Auto-contouring systems fall into three general types: commercial, open-source and developed in-house (Table 2). At the time of writing, there are nine CE-marked commercial solutions that can be used in the NHS to aid contouring for radiotherapy while more evidence is generated on assessing these technologies.[21]

The following points should be considered when a department is choosing an auto-contouring solution:

- Intended use, projected future applications and whether it will be used in research or clinical practice
- The data used to train the model and whether there is an option to use local data
- The regulatory position including CE or UKCA marking for medical devices and consideration of data protection requirements
- Costs including capital and ongoing operational costs
- Available support and documentation
- Staff skills required for commissioning and ongoing QA
- Whether it is compatible with current workflows, systems and network infrastructure.

03
Selection of system

When calculating the potential cost and time savings with an auto-contouring system, the time required by clinicians to both check and edit auto-contours should always be considered since the radiotherapy planning workflow may change.

Commercial solutions are costly and usually have a charge per case, which may include software licence and subscription, hardware, data storage, future upgrade and maintenance costs.[22] These products come with a ready-trained model library; however, the models may not suit the needs of a particular department and they provide limited opportunity to train local models. Pre-purchase assessments should therefore ideally be performed on local data sets. Companies will provide maintenance and technical documentation allowing ease of commissioning and ongoing quality control. The method of system installation is also important to consider, whether that be 'on premises' or via a cloud-based system. These different methods may have implications for potential loss of service if the manufacturer were to go out of business. Data security should always be considered for both local and cloud-based systems.

While open-source auto-contouring systems are ostensibly cheaper with no purchase costs, they are not CE marked as a medical device. Therefore, they require technical departmental expertise to set up and operate and may need to be taken through appropriate regulatory approvals before they can be used clinically. There is unlikely to be a maintenance contract so the department will need a long-term support plan. Some companies do provide commercial support of open-source products, and informal support may be available from the open-source developer community. Open-source solutions may come with pre-trained libraries and may also provide the ability to train bespoke models.

Finally, departments can develop their own in-house auto-contouring tool.[22] This requires adequate, well-curated training data and technically skilled staff. Collecting sufficient training data to develop an in-house system may take significant time. There is also a risk of failure if key staff leave, and therefore it is important that there is appropriate documentation on the model training, implementation, commissioning and QA. The Institute of Physics and Engineering in Medicine (IPEM) has guidelines on best practice for in-house manufacture of medical devices, including software, for use within the same institution.[23] For open-source software, the solution will not be CE marked and the department will need to be aware of the regulatory processes required to use this software clinically. Solutions used purely for research do not require CE marking or NICE approval.

03
Selection of system

Table 2. Comparison of the potential advantages (✓) and disadvantages (✗) of commercial, open-source and in-house auto-contouring solutions

		Commercial	Open-source	In-house
Capital costs	✓	--	May be free or cheaper than commercial system	May be free or cheaper than commercial system
	✗	Expensive	Costs of technically skilled staff	Costs of technically skilled staff
Operational costs	✓	Limited staffing costs	No maintenance contract	No maintenance contract
	✗	Contouring as a service has ongoing costs	Costs of technically skilled staff	Costs of technically skilled staff
Staff skills	✓	Limited technical skills required outside of commissioning	Opportunity to contribute to tool development	Interesting technical project
	✗	--	Relies on having technically skilled staff Risk of single point of failure if key staff leave	Relies on having technically skilled staff Risk of single point of failure if key staff leave
Regulatory position	✓	CE marked	--	Can be approved as an in-house development for clinical use
	✗	--	Regulatory requirements (not CE marked, software as a medical device)	Regulatory requirements (not CE marked, software as a medical device)
Support	✓	Extensive support and documentation available	Possible open-source development forums Can be well documented	--
	✗	Support and maintenance costs	May need to troubleshoot in-house	No support Documentation needs to be developed
Model training	✓	Ready-trained model library	May come with ready-trained models Opportunities to train own models	Full control of training data Knowledge of training data provenance
	✗	Models may not suit needs Unknown training data provenance Limited opportunity to train own models	Provenance of training data may or may not be known	Curation of suitable training data can be resource intensive (for multiple professions)
Ease of implementation	✓	Commissioning and ongoing QA only	Know how the system works	Know how the system works
	✗	'Black box'	Needs technically skilled staff	Needs technically skilled staff

04 Initial clinical evaluation and commissioning

- **Commissioning should ensure the safety and effectiveness of the auto-contouring system. This includes both the auto-contouring technology itself and the way it interacts with other elements of the radiotherapy treatment planning workflow.**
- **A multidisciplinary team should be involved in the design, management and assessment of commissioning tests.**
- **Commissioning should be performed on local data, representative of the real-world clinical cohort, including the full range of cases that will be encountered clinically.**
- **A combination of qualitative and quantitative methods of assessment is recommended. Geometric, dosimetric and time-based assessments should be coupled with a visual evaluation of the generated contours.**
- **The planning target volume (PTV) and planning organ at risk volume (PRV) margins should be considered for auto-contoured structures as there may be different uncertainties compared with manual contours.**

The purpose of commissioning is to ensure that the auto-contouring system is both safe and effective in the radiotherapy planning pathway. This includes the auto-contouring technology itself, the way it interacts with other systems and how it is managed within the radiotherapy treatment planning workflow for the intended use. A commercial auto-contouring system may perform differently between centres due to subtle differences in patients, contouring practices, imaging and radiotherapy planning pathways. It is therefore essential that these systems are appropriately validated in each department prior to use. The suggested mandatory and optimal commissioning requirements to ensure the safety and effectiveness of auto-contouring within a local radiotherapy department are summarised in Table 3.

Commissioning should be a multidisciplinary process, including but not limited to clinical oncologists, physicists, dosimetrists, therapeutic radiographers, IT and information governance. A range of staff groups should be involved in the design, management and assessment of commissioning tests and their results. Close collaboration with regulatory staff by the specialist project team is required. A robust approach to defining scope of use and providing users with detailed and specific information about auto-contouring performance and known failure modes and exclusions is critical to ensuring safety.

04
Evaluation and commissioning

In order to establish safety, it is necessary to ensure that the risk of clinically significant contouring errors is as low as reasonably achievable (ALARA) in line with the principles of radiation protection (IR(ME)R 2017).[3] It is important to distinguish between two types of significant error. Firstly, there may be a generally **poor segmentation performance** for a particular region of interest (ROI). This type of error will typically be easier to detect at commissioning as it will be common and affect most cases. Secondly, there may be **failure cases**, where a contour is clinically incorrect on an individual case, which can be rare (across the cohort) and/or local (within the scan set). Rare and local errors can be particularly challenging to detect in a limited cohort test at commissioning. Commissioning should be performed on local data, representative of the real-world clinical cohort, including the full range of cases that could be encountered clinically. Ideally, at least 20 local cases per clinical protocol should be used for commissioning testing. This will assess variation of performance across cases, particularly if some with significant anatomic variations (edge cases) are included. These edge cases may include patients with very high or low body size, prior surgery, metalwork, artifact or lack of contrast. Visual assessment of contours remains the primary method for detecting rare and local failure cases and this should be remembered throughout the commissioning and subsequent ongoing QA processes.

Commissioning tests

The types of assessment metrics that can be used to evaluate auto-contours are described in Section 2. When commissioning, it is important to undertake tests that evaluate how the system performs on local data, and there should be a combination of geometric, dosimetric and qualitative testing. Some of the important factors to check during commissioning are summarised in Table 4.

Geometric tests should be performed comparing auto-contours with 'gold-standard' contours. These could be a consensus contour, a simultaneous truth and performance level estimate (STAPLE) or quality-assured clinical contours evaluated by a minimum of two independent clinicians.[24,25] Various geometric tests are available as shown in Table 1. No single metric is universally accepted as reliable for detecting errors in all scenarios. Importantly, there is also no consensus on the thresholds for clinical acceptability for these metrics. Therefore, a suite of tests is preferred when commissioning a system, including overlap, distance and volume-based assessments. Rare, local but clinically significant geometric errors (failure cases) that occur in large organs are unlikely to be identified with any of these per-organ metrics. While maximal or near-maximal distance metrics could, in principle, detect such errors, they are often confounded in practice by noise from clinically insignificant differences (eg inclusion of distal bronchial tree in lung contour).

The clinical significance of contouring errors is difficult to determine from geometric measures alone, so dosimetric evaluation is an important part of commissioning. Plans should be produced using the auto-contours to evaluate the uncertainty this introduces into the planning process in comparison with manual 'gold-standard' contours. Since target prescription doses and OAR dose-volume constraints vary from site to site, and between clinical protocols (dose and fractionation), the dosimetric consequences of a given geometric error will vary. Moreover, as target sites are commonly variable in location, shape and size, a reasonable sample of patients is required to establish dosimetric and clinical consequences of auto-contouring errors. This sample size is likely to be larger than the sample on which geometric testing is performed. There may be exceptions to this where the tumour location is consistent, such as the clinical target volume for the prostate.

04
Evaluation and commissioning

Qualitative visual assessment is required as part of healthcare professional involvement in the assessment of auto-contours. The use of a qualitative assessment Likert scale may provide a structure for assessment during commissioning, which can be repeated during post-implementation monitoring to identify performance drift. There are currently multiple Likert scales in use for assessing auto-contours, and departments should consider carefully which scale they will use. Qualitative assessment can unfortunately be subjective, resulting in inter- and intra-rater variation in contour assessment. To minimise this, contours should be assessed in the context of the local contouring protocol and by the relevant clinical team.

Scope of use

Importantly, a scope of use for the auto-contouring system should be defined based on the acceptability of these test results. This should describe not only the contours that can be automatically generated, but also the range of imaging protocols, clinical protocols, dose and fractionation schedules and patient groups for which the auto-contouring has been commissioned. It should also make clear any exclusion criteria. Examples may include artificial hips, post-surgical cases, presence of pacemakers or extended field-of-view scans. Testing should include cases for these scenarios, and exclusion may be based on observed poor performance or higher rates of failure cases.

If auto-contours are to be used for target volumes or for OARs where a PRV is created, the margin required to create PTVs and PRVs should be considered. This should be done during the commissioning stage. Guidance is provided by the British Institute of Radiology, by considering the delineation uncertainty for auto-contoured structures as opposed to manually delineated structures.[21]

Effectiveness

Commissioning should consider how effective the overall system will be in delivering the desired benefits. These may be clinical (consistency, precision, ability to increase the number of contoured structures) or system (efficiency and time saving) benefits. For both, but particularly the latter, the way in which human operators interact with the system is crucial. Human users are busy, prone to automation bias, used to the types of errors other humans make and potentially sceptical of auto-contouring technologies. If implementation of auto-contouring systems is aimed primarily at time saving, departments should ensure that the reviewing and correcting of auto-contours does not take longer than manual contouring.

Importantly, the types of errors made by auto-contouring algorithms may be qualitatively different to those that human users make (see Table 5). Common errors should be detected during the commissioning process. Centres should collate a list of detected errors to enable user training, and clinicians should then be vigilant for these during the checking process.

04

Evaluation and commissioning

Table 3. Suggested mandatory and optimal requirements for auto-contouring systems

Mandatory commissioning requirements	Optimal commissioning requirements
Appropriate combination of geometric, qualitative, dosimetric and time-based assessments	Geometric tests should include overlap, volume and distance to agreement metrics
Performed on each relevant tumour site based on clinical use and risk assessment	Dosimetric tests should include clinically relevant dose statistics
Uses a local data set external to model development of sufficient size (at least 20 cases per clinical protocol) that represents the local patient cohort and imaging data distribution	Use an external data set from another centre to validate the system
Sets the 'scope of use' including the structures to be auto-contoured, any exclusion criteria and special conditions for use of auto-contours; differences between any training protocols should be noted here	Consider performance compared with inter- and intra-observer variation in manual contouring
Human visual qualitative assessment for clinically significant issues not detected by other testing	Consider impact on margins by estimating delineation uncertainty for auto-contouring systems
Collate a list of common errors detected during commissioning to enable user training	

05 Ongoing quality assurance: day-to-day use of auto-contours

- **Clinicians with responsibility for producing or approving manual contours according to local IR(ME)R protocols have ultimate responsibility for any AI-generated contours. A process must be in place for checking and editing these structures for every patient at an appropriate stage in their treatment.**
- **The reviewer must be familiar with the contouring protocol, and the auto-contouring system must be applicable to the clinical intent, indication and anatomical site being treated.**
- **The reviewer should use a structured approach for checking contours, and areas of critical dosimetric importance should be prioritised.**
- **More time should be allocated to review auto-contoured outputs in situations for which the software is known to underperform.**
- **Auto-contours should be subjected to usual peer review processes.**
- **Clinicians should be aware of the risks of automation bias and maintain a critical approach to evaluating auto-contours.**

Every automatically generated contour should be carefully reviewed, corrected if necessary and approved by clinical staff, with the clinical oncologist taking ultimate responsibility.[14] This should ideally take place before the treatment is planned. OAR contouring should undergo the same robust checking process as target volume delineation, with appropriate QA processes, particularly for OARs that are more difficult to define. A structured, systematic approach is required to perform these checks and, therefore, dedicated time should be allocated for review.[26] Review of all accepted auto-contoured targets and OARs at peer review meetings is encouraged.[27]

Contouring protocols

Inconsistency in contouring guidance may increase manual contour variability. Consistency and accuracy in structure nomenclature and contouring guidance not only minimises variation but also improves departmental workflow and safety, with positive impact on clinician peer review.[27] There should be standardised contouring and naming protocols for each anatomical site and indication to avoid inter-reviewer variability.

OARs should be contoured and labelled according to the Global Quality Assurance of Radiation Therapy Clinical Trials Harmonization Group (GHG) consensus guidelines as recommended in the RCR peer review guidelines.[27] If a target structure (eg breast, prostate, lymph node region) is auto-contoured, it should be carefully checked according to the relevant contouring protocol.

It is important that all nuances within a contouring protocol are well understood and followed in both the training of the AI software and checking of its output. Factors to consider in the commissioning, scope of use and daily evaluation of auto-contours are summarised in Table 4.

05
Quality assurance

Treatment intent and technique

Understanding clinical intent and treatment technique is crucial to evaluating clinical acceptability of the auto-contours. There may be different approaches for radical versus palliative treatment. Manual editing of an auto-contour should only be required if this will impact the treatment plan.

Where dose gradients are very steep and planning is pushed to OAR tolerance, delineation needs to be very precise. If a contouring error does not align with a steep dose gradient, it is less likely to be clinically significant. Reviewers can then focus efforts on the regions where they do coincide, and therefore are more likely to detect potential failure cases, where failure represents suboptimal patient treatment rather than contour deviation. With stereotactic radiotherapy, millimetre accuracy is required due to the high doses per fraction delivered over very few fractions. Small deviations in contours can significantly impact on treatment planning with stereotactic ablative radiotherapy (SABR) and brachytherapy as the dose-volume constraints include very small absolute OAR volumes. The use of daily online image-guided radiotherapy enables reduction of the CTV–PTV margins, and therefore the traditionally large margins can no longer be relied upon to protect against contouring deviations.

Table 4. Factors to consider in the commissioning, scope of use and daily evaluation of auto-contours

Consideration	Example
Patient cohort Is it the same?	• Size: is the system trained with patients from a different country or demographic? • Age: adult versus paediatric • Gender • Definitive treatment versus postoperative treatment • Implanted devices
Patient preparation Is it the same?	• Bladder and rectal status • Use of intravenous and/or oral contrast
Positioning and immobilisation Is it the same?	• Head first versus feet first • Prone versus supine position • Use of abdominal compression
Imaging modality Has the same modality been used for training the system? Imaging acquisition protocols, scanning parameters and image quality can affect AI tool performance	• T2-weighted magnetic resonance imaging (MRI) is used for spinal SABR whereas computed tomography (CT) is used for spinal cord contours in conventional radiotherapy (RT) oesophageal protocols (eg SCOPE-2 trial).[28] • Cross-modality should not be used. For example, when using an MRI-only workflow for prostate cancer treatment, both target and OARs should be trained and contours checked on MRI data sets and not applied from those trained on CT data sets.[29]

05
Quality assurance

Windowing Inappropriate windowing can result in large variations from the intended structure, both in size and inclusion/exclusion of anatomical features Ideally use preset window levels	• Lungs should be checked using **lung windowing** so that exclusion of target, hilar structures and bronchus are adequate. • Bowel loops should be defined using **abdominal windowing** to ensure the full extent of the bowel wall is included and gas contained within the organ is not mistaken for abdominal fat.
Clinical indication OAR definitions (and associated nomenclature) may differ across clinical indications Standardised nomenclature is being introduced across the UK	• Currently **heart alone** is contoured for oesophageal or lymphoma conventional RT whereas **heart and pulmonary artery** is required for lung SABR. • **Spinal cord** contoured on T2-weighted planning MRI is used for spine SABR, whereas all non-vertebral SABR indications currently use **spinal canal** contoured to bony limits.
Anatomical definition for dose-volume constraints Dose-volume constraints will have been derived from specific anatomical definitions of an organ	• For cervical cancer, **abdominal cavity** was used in the INTERLACE trial for the external beam RT, whereas **bowel loops** were used in the EMBRACE II study.[30,31] • In addition, pelvic brachytherapy and SABR indications both use bowel loops.[32]
Clinical intent Can determine acceptability	• Radical versus palliative treatment.
Treatment technique Will intensity-modulated radiation therapy (IMRT) or volumetric modulated arc therapy (VMAT), SABR or brachytherapy be used? Is online adaptive radiotherapy to be used?	• Dose and OAR tolerances may be specific to technique. • Use of absolute versus percentage volume of OAR for dose tolerances may impact on the dosimetric impact of an error.

Checking process

Auto-contouring systems often produce structures that have smooth edges in all directions and look an appropriate shape. This can make some errors difficult to detect and can introduce **anchoring bias**, where the reviewer is inappropriately influenced by the contour that is already present. When reviewing auto-contours, clinicians need to be mindful that auto-contouring errors are different to manual errors. The auto-contouring system may also fail on individual cases in ways that have not been detected in commissioning. The day-to-day checking process should detect these errors, but close attention and a systematic approach are required. The clinician should also be mindful of the time taken to edit a structure compared with the time taken to contour from scratch, if a contour needs significant alterations.

05
Quality assurance

Clinicians should check each auto-contoured structure in turn, prioritising those with the greatest dosimetric importance, which are the target volumes and the OARs in close proximity to the PTV. Each structure should be checked in each plane (axial, sagittal and coronal) on each slice using a systematic approach. This is important to verify completeness and consistency of delineation in all dimensions.[33] The review should be performed on the appropriate imaging modality and window setting for the OAR and clinical indication. Generation and use of checklists, possibly by anatomical site, may help to reduce the risk of automation bias during the review process and to document that human review has occurred. Applying the Likert scale used for qualitative assessment in commissioning may also provide a standardised, structured approached to determining if the auto-contoured structures need editing or redrawing from scratch. When reviewing a set of contours all factors affecting the definition of those structures, as described in the previous section, should be carefully considered and applied to the review. In addition, any additional volumes required, such as PRVs, should be created or modifications made to the contours, such as editing OARs that overlap with the target.

Generalised common errors for each structure may be detected at commissioning and these should be specially looked for in each day-to-day review. Image co-registration inaccuracies and artifacts affecting image quality may impact contouring accuracy and precision; clinical oncologists should be aware of these potential sources of error and review the final contours on the primary data set.[28] Under-contouring of the OAR may lead to inferior OAR sparing with potential for increased or unanticipated toxicity, and over-contouring could result in unnecessary dose compromises to the target. In view of the growing use of sequential and multimodality anti-cancer therapies, inaccuracies in OAR contouring, and hence plan optimisation, risk inappropriate dose delivery to an OAR, with greater potential for 'dose-dumping' in normal tissues and subsequent unanticipated toxicity during a patient's treatment pathway.[28,34]

Departments should consider additional checking processes to ensure that appropriate review of auto-contours has taken place. This may include renaming of the auto-contours, independent contour review and generation of volume metrics comparing the auto-contour with the structure approved for clinical use.

Automation bias

Automation bias is an important source of error in auto-contouring for radiotherapy. This is a phenomenon whereby reviewers eventually favour output from the automated system, despite having evidence or knowledge that would suggest the automated system is wrong.[11] Contouring errors can be one of the biggest sources of systematic error in radiotherapy so it is imperative all auto-contours are checked appropriately.[21] Clinicians need to be aware of the risks of automation bias when reviewing auto-contours.[35] Suggestions to reduce the risk of automation bias include educating users about AI reasoning processes, emphasising human accountability, presenting AI output uncertainties and providing training on the reliability of specific situations.[19] Auto-contouring systems make different errors to humans and it is important that clinician-reviewers are aware of this. The use of checklists and standardised processes has also been shown to reduce the risk of automation bias and increase safety.[36]

05
Quality assurance

Common errors

The errors in AI-generated structures are often very different to those arising from manual contouring. There are various situations in which auto-contouring is known to underperform and these should be assessed more diligently, with more time allocated for the review. Some examples are listed in Table 5. When an auto-contouring error is identified this should be recorded to ascertain if this is a persistent variation and possibly highlight it within the radiotherapy pathway. This may also be fed back to manufacturers.

Table 5. Common types of auto-contouring errors

Type of error	Examples

Contrast density	• Inclusion of distant organs with similar contrast density. These may be resolved with post-processing.
	• Omission of bowel gas and/or the bowel wall adjacent to bowel gas.
	• Inclusion of adjacent air-containing structures:
	  ◦ Trachea
	  ◦ Oesophagus

Structure length	• Over- and under-contouring of structures whereby only a subsection of the organ is contoured according to the protocol:
	  ◦ Spinal cord, oesophagus, femoral heads
	• Boundaries of adjacent structures should be in continuity and not overlap:
	  ◦ Brainstem and spinal cord
	  ◦ Stomach and duodenum
	• Boundaries of OAR may incorrectly overlap with tumour or target volume.

Distortion by tumour	• OAR in close proximity to a tumour may be displaced, deformed and/or invaded by the tumour: ◦ Oesophagus next to a mediastinal tumour
Disturbance by artifact	• Auto-contouring system may not recognise metallic implants (eg hip replacements and dental work), which may result in atypical anatomy but also cause image artifacts that may impact system performance.

Atypical anatomy	• Auto-contouring system may not be familiar with abnormal anatomy:
	  ◦ Horseshoe kidney, previous bowel resection, previous hysterectomy, scoliosis
	  ◦ Abdominal compression may distort abdominal structures
	• Auto-contouring system may not recognise different treatment set-up positions or population cohorts (eg paediatric).

Processing problems	• Missing slices • Holes in structures • Disconnected components in structures and irregular edges
Contour deviations on small volumes	• Small deviations can have significant errors with small target volumes or OARs: ◦ Cochlea, penile bulb, optic chiasm, optic nerve

06 Post-implementation monitoring

- **Ongoing surveillance of the performance and clinical utility of the auto-contouring technology after clinical implementation must be undertaken.**
- **Performance can change with time due to changes in input imaging, changes in treatment indication that move further from the initial training data and changes in the patient population. When performance drift is identified, this should be fed back to the model manufacturer.**
- **Careful consideration should be given to system changes that might cause a drift in overall performance. It is important that auto-contouring systems can be easily tracked over the entire life cycle of the product.**
- **The clinical use of auto-contouring systems may lead to subtle changes in user perspective. The effects of anchoring bias and review fatigue may result in less identification of edge cases or poor system performance.**
- **Regular formal post-implementation evaluation of models against a test library should be performed.**

While the majority of assessment activity will occur when an auto-contouring system is initially commissioned for clinical use, it should be considered that there are likely to be changes in the input imaging data sets during the system life cycle. Changes in clinical criteria for treatment will result in changes in patient demographic and disease burden, both of which can impact on body habitus. Organ separation by fat is known to impact auto-contouring system performance. While a new CT simulator is likely to be evaluated for image quality at the time of commissioning, software upgrades that include changes in reconstruction algorithm may be introduced without consideration of the effect on auto-contouring systems.

Where an auto-contouring system makes use of a trainable model, it is generally understood that a retrained model should be labelled with a new version number and undergo a proportionate evaluation prior to clinical use, as would be performed for an updated radiotherapy treatment planning system. It should also be considered that the underlying models of auto-contouring systems are not fully deterministic in nature, such that a change in the underlying hardware, software library or available memory may result in different performance of the same model. The effect of such ‘drift’ may be difficult to detect on an individual patient basis, specifically due to the joint effects of anchoring bias and review fatigue. Review fatigue refers to the fact that a systematic error of an auto-contouring algorithm that becomes gradually more prevalent may not be escalated by a user as they become increasingly familiar with the auto-contouring system.

06
Monitoring

Best practice to mitigate against these risks is to perform a scheduled re-evaluation of the auto-contouring system against a local library of test image data sets, for which the performance of the model has already been assessed, and to gradually add to the cases in the library over time. Evaluations should include quantitative geometric assessments, ideally automated, and may also include brief human-centric evaluation of contour quality. Performing formal regular assessments at least once a year and with any change in imaging system should identify if there is any drift in performance.

The system of ongoing QA outlined above should be apparent in the data stored at an individual patient level. The digital imaging and communications in medicine (DICOM) imaging standard offers a mechanism to encode information about software systems and versions that were used to generate secondary image data, such as contouring information. Most manufacturers will encode system and version information into the DICOM headers of output files. However, as DICOM data are moved between radiotherapy systems, the output of a software system may reconstruct a new DICOM image without preserving all previous information. This has the effect of breaking the 'audit trail' in the image data and making it harder to track the performance of specific versions of the auto-contouring system over time. The simplest method to mitigate against this risk is to ensure that the original DICOM image objects created by the auto-contouring system are stored in a post-treatment archive.

Other post-monitoring approaches may include structured feedback loops to allow clinicians and radiographers to report discrepancies in a standardised manner, and undertaking regular review meetings to identify areas for improved real-world usage.

07 Impact on the multidisciplinary team

- **The radiotherapy planning workflow involves a multidisciplinary team with specific roles at each step of the pathway. Dedicated time to review and edit auto-contours should be appropriately built into this pathway.**
- **Job plans may need to be updated to account for changes in workflow.**
- **Training the multidisciplinary team members to use and assess auto-contours should focus on both safety and efficiency. Team members undertaking QA should be trained to understand that auto-contouring errors may be qualitatively different to human errors.**
- **Departments should be clear about who is taking overall responsibility for the checking and editing of auto-contours. Roles detailed in IR(ME)R 2017 are unchanged by the use of auto-contouring systems.**
- **A fully manual pathway should always be available in case of software failure or clinical trial specification.**

Consideration is needed of the impact on the wider radiotherapy planning pathway and changes in roles within the multidisciplinary team. Although clinical oncologists are ultimately responsible for the final treatment plan, many radiotherapy departments routinely delegate the contouring and QA of OARs to trained planning staff to free up clinician time.

It is important to ensure workflow efficiency gains in one place are not lost elsewhere. Examples of this could include an increase in replans due to contouring errors picked up at a physics check or subsequent peer review, or a tendency to unnecessarily edit auto-contours to make them appear more like human-derived contours, while having minimal impact on radiotherapy planning. The additional workload associated with commissioning and the ongoing QA should also be considered.

During the introduction of an auto-contouring pathway, departments will need to consider the wider team competence and responsibility for reviewing and correcting auto-contoured structures. Job plans may also need to be reviewed if the radiotherapy planning workflow is to change.[37] It must be emphasised that auto-contouring is a tool that can support existing workflows, rather than replace jobs. It can support the introduction of more complex techniques and therefore may not necessarily reduce clinician contouring times. Protected review time is essential as detection of errors requires careful evaluation with a systematic approach for each case. In the future, auto-contouring may also permit an increase in online adaptive radiotherapy, which will require further changes to the team structure and workflow.

08 Education and training

- **All healthcare professionals with responsibility for reviewing auto-contours must be able to perform and maintain their skills in manual delineation. They must understand the relevant anatomy and contouring protocol to safely check auto-contours.**
- **Reviewing auto-contours is a different skill to manual delineation and this requires additional training.**
- **There should be a robust education and training structure to ensure awareness of:**
  - **The process of clinical implementation**
  - **The limitations of the available auto-contouring systems on local patient populations and protocols, with a framework for continued peer learning**
  - **Where to gather learning and information from the specialist implementation team.**
- **The scope of auto-contouring is likely to evolve and further educational resources will be required for continuing professional development (CPD).**

All healthcare professionals approving auto-contours will require sufficient knowledge about auto-contouring, including how to perform day-to-day QA as highlighted in Section 5. They should also be aware of the limitations of auto-contouring and the risks of automation bias. These healthcare professionals will need education and training to be able to achieve this knowledge. The scope of auto-contouring is also likely to change, and these healthcare professionals should stay up to date as part of CPD.

Clinicians with an educational or leadership responsibility should be aware of the need to support the education of both existing healthcare professionals and those in training across the whole multidisciplinary team. Departments may wish to provide formal training in auto-contouring implementation or provide mechanisms for clinicians to undertake self-study. Simulation-based training may be helpful to reflect the challenges in auto-contouring technology. All healthcare professionals using auto-contouring should know where to seek further information. Auto-contouring is likely to affect how training in radiotherapy is delivered and this should be considered when producing relevant curricula.

Reviewing and editing an auto-contour is a separate skill to performing manual segmentation. All healthcare professionals involved in radiotherapy delineation must be experienced in manual segmentation with a structured mandatory training programme. This is to allow safe continuation of the radiotherapy treatment pathway for non-standard and complex cases, or in the event of loss or failure of auto-contouring pathways. Maintaining these skills in a highly automated environment may require clinicians to intentionally perform manual primary segmentation on a subset of cases. Radiotherapy planning pathways must provide opportunities for trainees to undertake manual segmentation of clinical cases to develop the next generation of experienced clinicians.

08

Education and training

For review and editing of auto-contours, clinical users require an understanding of the limitations of local auto-contouring models. This includes differences in the model training population and volume definitions versus local population and protocols, regions of poor performance and systematic errors identified during local commissioning, typical failure cases and understanding the risks of automation bias. Clinical cases with poor auto-contouring performance should be routinely discussed at peer review meetings to identify systematic errors and failure modes, especially on introduction of any new or updated models. Departments may consider having local training data sets of erroneous auto-contouring cases for new clinical users to review.

Clinical users should ensure CPD in their contouring competencies and that peer review remains embedded in auto-contouring workflows as per standards laid out in the RCR guidance.[27]

09 Research

- **The local team must check that clinical trial protocols permit the use of auto-contours. When permitted, the system must be compliant with the latest version of the clinical trial protocol. If the system is not compliant, manual contouring should be performed.**
- **Auto-contouring software developed for research purposes must be subjected to regulatory processes before being used clinically.**
- **Auto-contour uncertainties may impact on study results, and researchers should incorporate these uncertainties in their analysis plan.**

Centres must be aware of any differences between the contouring protocols used for the AI training data set and protocols that are being followed in clinical trials. In addition, clinical trials may explore different patient cohorts and indications than those used to commission the auto-contouring system. Auto-contours must always be edited to ensure agreement with clinical trial protocol definitions. Researchers should ensure the trial protocol is permissive of the use of AI-assisted contouring.

Clinical trials and other ongoing research may lead to changes in target volume or OAR definitions and can introduce new OARs or critical subvolumes that should be considered (eg cardiac substructures).[38] Data sets may not be available in these situations to train the model in the first instance, so special attention must be given to differences from the new standard of care. It may be necessary to revert to manual contouring until these changes have been implemented by the AI system manufacturers.

Auto-contouring may be used in treatment stratification strategies. Personalised medicine and decision-making are likely to become areas where auto-contouring may also contribute to faster uptake of research outcomes.

It is important to understand the wider impact of auto-contouring on other types of research. There is vast scope for use of auto-contouring in data mining and big data analyses for generation of 'real-world' evidence and application in the field of radiomics, which is expected to play a vital role in both tumour detection and diagnosis as well as prediction of treatment outcomes.[39,40] Many image-based biomarker studies rely on defining structures in order to calculate the features of the biomarker. It is well established, for example, that variability in contouring is a major cause of uncertainty in radiomics analyses, but the theoretically improved consistency of auto-contouring may consequently lead to improved consistency in radiomics signature definition.[41,42] Auto-contouring may also be employed for service evaluation or service improvement. Retrospective review of dose to various normal tissues, whether defined as at risk or not, can lead to improved planning techniques and correlation with toxicity.

In any research setting, it is important to note that poor contouring will result in poor results. It is therefore critical that contouring is accurate, precise and consistent.

10 Additional considerations

- **The scope for auto-contouring systems will continue to evolve. Any new applications must be commissioned, validated and follow a rigorous ongoing QA process.**
- **Auto-contouring is likely to impact on many aspects of radiotherapy treatment delivery in the future including, but not limited to, the radiotherapy workflow, education, job planning and research. Future workforce and job planning will need to take this into consideration.**
- **Centres should consider having information available for patients on the use of artificial intelligence in their treatment pathway.**

The scope for auto-contouring systems will continue to evolve to include tumour target volumes, more tumour types and different imaging modalities. Commercial systems do not offer gross tumour volume (GTV) contouring for target delineation at present, but this is an area of ongoing research in various anatomical sites using different imaging modalities.[43,44,45] This will require rigorous commissioning and checking if it becomes available clinically.

Auto-contouring tools have great potential to support online adaptive radiotherapy (oART).[46,47] Currently many systems employ structure propagation and manual contouring.[48,49,50] The quality and speed of online imaging, depending on the treatment platform and imaging modality, will dictate the adoption of AI in this setting. It is important the same recommendations are followed for online planning as when using AI in the initial treatment planning process. It should be noted that the auto-contouring tool will need to be trained with on-set imaging data sets and not those used for initial planning. Time restrictions will apply in the oART setting and therefore careful attention to efficiency of the process is critical, with availability of appropriately trained staff being essential.

With increasing adoption of AI technologies in radiotherapy, departments should consider how to reassure patients on the use of AI in their treatment pathway. They should therefore have information available for patients, should they request it, which could include explanation of the workflows, implementation and checking processes with human supervision.[51,52,53,54,55]

References

1 Hindocha S, Zucker K, Jena R *et al*. Artificial intelligence for radiotherapy auto-contouring: current use, perceptions of and barriers to implementation. *Clin Oncol (R Coll Radiol)* 2023; **35**: 219–26. 10.1016/j.clon.2023.01.014.

2 National Institute for Health and Care Excellence (NICE). Artificial intelligence technologies to aid contouring for radiotherapy treatment planning: early value assessment. NICE, 2023. www.nice.org.uk/guidance/hte11/chapter/3-Committee-discussion (accessed 16/11/23).

3 The Ionising Radiation (Medical Exposure) Regulations 2017.

4 Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. *Semin Radiat Oncol* 2019; **29**: 185–97. 10.1016/j.semradonc.2019.02.001.

5 Montague E, Roques T, Spencer K, Burnett A, Lourenco J, Thorp N. How long does contouring really take? Results of the Royal College of Radiologists contouring surveys. *Clin Oncol (R Coll Radiol)* 2024; **36**: 335–42. 10.1016/j.clon.2024.03.005.

6 Medicines and Healthcare products Regulatory Agency (MHRA). Crafting an intended purpose in the context of software as a medical device (SaMD). MHRA, 2023. www.gov.uk/government/publications/crafting-an-intended-purpose-in-the-context-of-software-as-a-medical-device-samd/crafting-an-intended-purpose-in-the-context-of-software-as-a-medical-device-samd (accessed 09/05/2024).

7 Medicines and Healthcare products Regulatory Agency (MHRA). Medical devices: software applications (apps). MHRA, 2023. https://assets.publishing.service.gov.uk/media/64a7d22d7a4c230013bba33c/Medical_device_stand-alone_software_including_apps__including_IVDMDs_.pdf (accessed 09/05/24).

8 Regulatory Horizons Council. *The regulation of artificial intelligence as a medical device*. Regulatory Horizons Council, 2022. https://assets.publishing.service.gov.uk/media/6384bf98e90e0778a46ce99f/RHC_regulation_of_AI_as_a_Medical_Device_report.pdf (accessed 09/05/24).

9 NHS England. Digital Clinical Safety Informatics Team. NHS England, 2024. https://digital.nhs.uk/services/clinical-safety (accessed 09/05/24).

10 Brouwer CL, Dinkla AM, Vandewinckele L *et al*. Machine learning applications in radiation oncology: current use and needs to support clinical implementation. *Phys Imaging Radiat Oncol* 2020; **16**: 144–8. 10.1016/j.phro.2020.11.002.

11 Nealon KA, Balter PA, Douglas RJ *et al*. Using failure mode and effects analysis to evaluate risk in the clinical adoption of automated contouring and treatment planning tools. *Pract Radiat Oncol* 2022; **12**: e344-e53. 10.1016/j.prro.2022.01.003.

12 Harrison K, Pullen H, Welsh C, Oktay O, Alvarez-Valle J, Jena R. Machine learning for auto-segmentation in radiotherapy planning. *Clin Oncol (R Coll Radiol)* 2022; **34**: 74–88. 10.1016/j.clon.2021.12.003.

13 Chen M, Wu S, Zhao W, Zhou Y, Zhou Y, Wang G. Application of deep learning to auto-delineation of target volumes and organs at risk in radiotherapy. *Cancer Radiother* 2022; **26**: 494–501. 10.1016/j.canrad.2021.08.020.

14 Vandewinckele L, Claessens M, Dinkla A *et al*. Overview of artificial intelligence-based applications in radiotherapy: recommendations for implementation and quality assurance. *Radiother Oncol* 2020; **153**: 55–66. 10.1016/j.radonc.2020.09.008.

15 Sherer MV, Lin D, Elguindi S *et al*. Metrics to evaluate the performance of auto-segmentation for radiation treatment planning: a critical review. *Radiother Oncol* 2021; **160**: 185–91. 10.1016/j.radonc.2021.05.003.

16 Vaassen F, Hazelaar C, Vaniqui A *et al*. Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy. *Phys Imaging Radiat Oncol* 2020; **13**: 1–6. 10.1016/j.phro.2019.12.001.

17 Nikolov S, Blackwell S, Zverovitch A *et al*. Clinically applicable segmentation of head and neck anatomy for radiotherapy: deep learning algorithm development and validation study. *J Med Internet Res* 2021; **23**: e26151. 10.2196/26151.

18 Mackay K, Bernstein D, Glocker B, Kamnitsas K, Taylor A. A review of the metrics used to assess auto-contouring systems in radiotherapy. *Clin Oncol (R Coll Radiol)* 2023; **35(6)**: 354–369. 10.1016/j.clon.2023.01.016.

19 McCague C, MacKay K, Welsh C, Constantinou A, Jena R, Crispin-Ortuzar M. Position statement on clinical evaluation of imaging AI. *Lancet Digit Health* 2023; **5**: e400-e2. 10.1016/s2589-7500(23)00090-0.

20 Tudor G, Bernstein D, Riley S *et al*. *Geometric uncertainties in daily online IGRT: refining the CTV–PTV margin for contemporary photon radiotherapy*. British Institute of Radiology, 2020. 10.1259/geo-unc-igrt.

21 National Institute for Health and Care Excellence (NICE). *Artificial intelligence technologies to aid contouring for radiotherapy treatment planning: early value assessment*. NICE, 2023. www.nice.org.uk/guidance/hte11 (accessed 10/05/24).

22 Marchant T, Price G, McWilliam A *et al*. Assessment of heart-substructures auto-contouring accuracy for application in heart-sparing radiotherapy for lung cancer. *BJR Open* 2024; **6**: tzae006. 10.1093/bjro/tzae006.

23 Institute of Physics and Engineering in Medicine (IPEM). *Guidance for health institutions on in-house manufacture and use, including software*. IPEM, 2022. https://www.ipem.ac.uk/media/yvdhwxw4/ipem-best-practiceguidance-on-ihmu-v2-2_final.pdf (accessed 26/09/2024).

24 Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 2004; **23**: 903–21. 10.1109/TMI.2004.828354.

25 Gwynne S, Spezi E, Sebag-Montefiore D *et al*. Improving radiotherapy quality assurance in clinical trials: assessment of target volume delineation of the pre-accrual benchmark case. *Br J Radiol* 2013; **86**: 20120398. 10.1259/bjr.20120398.

26 The Royal College of Radiologists (RCR). *Radiotherapy target volume definition and peer review*. Second edition. RCR guidance. London: The Royal College of Radiologists, 2022. www.rcr.ac.uk/publication/radiotherapy-target-volume-definition-and-peer-review-second-edition-rcr-guidance (accessed 24/11/22).

27 Mir R, Kelly SM, Xiao Y *et al*. Organ at risk delineation for radiation therapy clinical trials: Global Harmonization Group consensus guidelines. *Radiother Oncol* 2020; **150**: 30–9. 10.1016/j.radonc.2020.05.038.

28 Bridges S, Thomas B, Radhakrishna G *et al*. SCOPE 2: still answering the unanswered questions in oesophageal radiotherapy? SCOPE 2: a randomised phase II/III trial to study radiotherapy dose escalation in patients with oesophageal cancer treated with definitive chemoradiation with an embedded phase II trial for patients with a poor early response using positron emission tomography/computed tomography. *Clin Oncol (R Coll Radiol)* 2022; **34**: e269-e80. 10.1016/j.clon.2022.03.019.

29 Elguindi S, Zelefsky MJ, Jiang J *et al*. Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy. *Phys Imaging Radiat Oncol* 2019; **12**: 80–6. 10.1016/j.phro.2019.11.006.

30 Eminowicz G, Hall-Craggs M, Diez P, McCormack M. Improving target volume delineation in intact cervical carcinoma: literature review and step-by-step pictorial atlas to aid contouring. *Pract Radiat Oncol* 2016; **6**: e203-e13. 10.1016/j.prro.2016.01.006.

31 Tanderup K, Pötter R, Lindegaard J *et al*. EMBRACE II study protocol v1.0: image guided intensity modulated external beam radiochemotherapy and MRI based adaptive BRAchytherapy in locally advanced CErvical cancer. 2015. www.embracestudy.dk/UserUpload/PublicDocuments/EMBRACE%20II%20Protocol.pdf (accessed 23/09/24).

32 Chmura S, Winter KA, Robinson C *et al*. Evaluation of safety of stereotactic body radiotherapy for the treatment of patients with multiple metastases: findings from the NRG-BR001 phase 1 trial. *JAMA Oncol* 2021; **7**: 845–52. 10.1001/jamaoncol.2021.0687.

33 Wright JL, Yom SS, Awan MJ *et al*. Standardizing normal tissue contouring for radiation therapy treatment planning: an ASTRO consensus paper. *Pract Radiat Oncol* 2019; **9**: 65–72. 10.1016/j.prro.2018.12.003.

34 Delaney AR, Dahele M, Slotman BJ, Verbakel W. Is accurate contouring of salivary and swallowing structures necessary to spare them in head and neck VMAT plans? *Radiother Oncol* 2018; **127**: 190–6. 10.1016/j.radonc.2018.03.012.

35 Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc* 2012; **19**: 121–7. 10.1136/amiajnl-2011-000089.

36 Fong de Los Santos LE, Evans S, Ford EC *et al*. Medical physics practice guideline 4.a: development, implementation, use and maintenance of safety checklists. *J Appl Clin Med Phys* 2015; **16**: 5431. 10.1120/jacmp.v16i3.5431.

37 The Royal College of Radiologists (RCR). *Clinical oncology job planning guidance for consultant and SAS doctors*. London: The Royal College of Radiologists, 2022. https://www.rcr.ac.uk/our-services/all-our-publications/clinical-oncology-publications/clinical-oncology-job-planning-guidance-for-consultant-and-sas-doctors-2022

38 Walls GM, McCann C, Ball P *et al*. A pulmonary vein atlas for radiotherapy planning. *Radiother Oncol* 2023; **184**: 109680. 10.1016/j.radonc.2023.109680.

39 Naser MA, Wahid KA, Grossberg AJ *et al*. Deep learning auto-segmentation of cervical skeletal muscle for sarcopenia analysis in patients with head and neck cancer. *Front Oncol* 2022; **12**: 930432. 10.3389/fonc.2022.930432.

40 Nie K, Al-Hallaq H, Li XA *et al*. NCTN assessment on current applications of radiomics in oncology. *Int J Radiat Oncol Biol Phys* 2019; **104**: 302–15. 10.1016/j.ijrobp.2019.01.087.

41 Fornacon-Wood I, Faivre-Finn C, O'Connor JPB, Price GJ. Radiomics as a personalized medicine tool in lung cancer: separating the hope from the hype. *Lung Cancer* 2020; **146**: 197–208. 10.1016/j.lungcan.2020.05.028.

42 Papanikolaou N, Matos C, Koh DM. How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging* 2020; **20**: 33. 10.1186/s40644-020-00311-4.

43 Liang Y, Schott D, Zhang Y *et al*. Auto-segmentation of pancreatic tumor in multi-parametric MRI using deep convolutional neural networks. *Radiother Oncol* 2020; **145**: 193–200. 10.1016/j.radonc.2020.01.021.

44 Gay SS, Cardenas CE, Nguyen C *et al*. Fully-automated, CT-only GTV contouring for palliative head and neck radiotherapy. *Sci Rep* 2023; **13**: 21797. 10.1038/s41598-023-48944-2.

45 Wei Z, Ren J, Korreman SS, Nijkamp J. Towards interactive deep-learning for tumour segmentation in head and neck cancer radiotherapy. *Phys Imaging Radiat Oncol* 2023; **25**: 100408. 10.1016/j.phro.2022.12.005.

46 Lamb J, Cao M, Kishan A *et al*. Online adaptive radiation therapy: implementation of a new process of care. *Cureus* 2017; **9**: e1618. 10.7759/cureus.1618.

47 Byrne M, Archibald-Heeren B, Hu Y *et al*. Varian ethos online adaptive radiotherapy for prostate cancer: early results of contouring accuracy, treatment plan quality, and treatment time. *J Appl Clin Med Phys* 2022; **23**: e13479. 10.1002/acm2.13479.

48 Henke L, Kashani R, Robinson C *et al*. Phase I trial of stereotactic MR-guided online adaptive radiation therapy (SMART) for the treatment of oligometastatic or unresectable primary malignancies of the abdomen. *Radiother Oncol* 2018; **126**: 519–26. 10.1016/j.radonc.2017.11.032.

49 Green OL, Henke LE, Hugo GD. Practical clinical workflows for online and offline adaptive radiation therapy. *Semin Radiat Oncol* 2019; **29**: 219–27. 10.1016/j.semradonc.2019.02.004.

50 Kurz C, Buizza G, Landry G *et al*. Medical physics challenges in clinical MR-guided radiotherapy. *Radiat Oncol* 2020; **15**: 93. 10.1186/s13014-020-01524-4.

51 Aung YYM, Wong DCS, Ting DSW. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. *Br Med Bull* 2021; **139**: 4–15. 10.1093/bmb/ldab016.

52 Haan M, Ongena YP, Hommes S, Kwee TC, Yakar D. A qualitative study to understand patient perspective on the use of artificial intelligence in radiology. *J Am Coll Radiol* 2019; **16**: 1416–9. 10.1016/j.jacr.2018.12.043.

53 Temple S, Rowbottom C, Simpson J. Patient views on the implementation of artificial intelligence in radiotherapy. *Radiography (Lond)* 2023; **29** Suppl 1: S112–s6. 10.1016/j.radi.2023.03.006.

54 Yakar D, Ongena YP, Kwee TC, Haan M. Do people favor artificial intelligence over physicians? A survey among the general population and their view on artificial intelligence in medicine. *Value Health* 2022; **25**: 374–81. 10.1016/j.jval.2021.09.004.

55 York T, Jenney H, Jones G. Clinician and computer: a study on patient perceptions of artificial intelligence in skeletal radiography. *BMJ Health Care Inform* 2020; **27**. 10.1136/bmjhci-2020-100233.

Authors

Lead author – Dr Katherine Mackay, Specialist Registrar in Clinical Oncology, Royal Marsden Hospital and NHS Fellow in Clinical AI

Lead author – Dr Alexandra Taylor, Consultant in Clinical Oncology, The Royal Marsden NHS Foundation Trust

Dr Kathryn Banfill, Consultant in Clinical Oncology, The Christie NHS Foundation Trust

Dr David Bernstein, Radiotherapy Physicist, The Royal Marsden NHS Foundation Trust

Dr Jim Daniel, Head of Treatment Planning and Brachytherapy, South Tees NHS Foundation Trust, The Institute in Physics & Engineering in Medicine and Radiotherapy SIG representative

Dr Patricia Diez, Lead Clinical Scientist, Mount Vernon Cancer Centre

Dr Sarah Gwynne, Consultant in Clinical Oncology, Southwest Wales Cancer Centre

Dr Andrew Hoole, Medical Physicist, Cambridge University Hospitals NHS Foundation Trust

Dr Rajesh Jena, Academic Consultant Neuro-oncologist, The Royal Marsden NHS Foundation Trust

Dr Thomas Marchant, Clinical Scientist, The Christie NHS Foundation Trust

Dr Michael Nix, Principal Clinical Scientist, Leeds Teaching Hospitals NHS Trust

Dr Gareth Price, Clinical Scientist, The Christie NHS Foundation Trust

Dr Mark Teo, Consultant in Clinical Oncology, Leeds Teaching Hospitals NHS Trust

Acknowledgements

The Royal College of Radiologists team

Ms Stephanie James, Professional Standards Manager, The Royal College of Radiologists

Dr Andrew Leary, AI Manager, The Royal College of Radiologists

Clinical Oncology AI Group

Dr Ian Boon, Consultant Clinical Oncologist, University Hospital Southampton NHS Foundation Trust

Dr Sumeet Hindocha, Consultant Clinical Oncologist, University College London Hospitals NHS Foundation Trust

Dr Imogen Locke, Consultant Clinical Oncologist, The Royal Marsden NHS Foundation Trust

Dr Delia Pudney, Consultant Clinical Oncologist, Swansea Bay University Health

Dr James Wang, Consultant in Clinical Oncology, University College London Hospitals NHS Foundation Trust

Dr Kieran Zucker, Clinical Lecturer in Health Data Science and Machine Learning Engineering and Honorary Clinical Oncology Specialist Registrar, University of Leeds

Auto-contouring in radiotherapy meeting attendees

Prof Neil Bolger, Clinical Scientist, Addenbrooke's Hospital

Dr Catherine Clark, Medical Physicist, University College London Hospitals NHS Foundation Trust

Prof Ben Glocker, Computer Scientist, Imperial College London

Spencer Goodman, Professional Officer for Radiotherapy and Cancer Care, Society and College of Radiographers

Dr Teresa Guerrero Urbano, Consultant Clinical Oncologist, Guy's and St Thomas NHS Foundation Trust

Dr Louise Hanna, Consultant Clinical Oncologist, Velindre Cancer Centre

Dr Sooha Kim, Clinical Oncology Registrar, Institute of Cancer Research

Dr Ami Mehta, National Clinical Fellow in Radiotherapy NHSE and the Royal College of Radiologists and Clinical Oncology Registrar, NHS England and the Royal College of Radiologists

Dr Rafael Moleron, Consultant Clinical Oncologist, NHS Grampian

Dr Meera Nandhabalan, Consultant Clinical Oncologist, Oxford University Hospitals

Prof Siram Padmanaban, Consultant Physicist, Oxford University Hospitals

Dr Samantha Warren, Consultant Clinical Scientist, The Newcastle Upon Tyne Hospital

Nicky Whilde, Head of Radiotherapy Physicist, Mid and South East Essex NHS Foundation Trust

The Royal College of Radiologists
63 Lincoln's Inn Fields
London, WC2A 3JW, UK

The Royal College of Radiologists
is a Charity registered with the
Charity Commission No 211540.

+44 020 7405 1282
enquiries@rcr.ac.uk
rcr.ac.uk

@RCRadiologists

The Royal College of Radiologists

The Royal College of Radiologists. *Guidance on auto-contouring in radiotherapy*. London: The Royal College of Radiologists, 2024.

The Royal College of Radiologists is a Charity registered with the Charity Commission No. 211540