NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

EARLY VALUE ASSESSMENT PROGRAMME

Early value assessment guidance consultation document

Artificial intelligence technologies to aid contouring for radiotherapy treatment planning: early value assessment

Guidance development process

Early value assessment (EVA) guidance rapidly provides recommendations on promising health technologies that have the potential to address national unmet need. NICE has assessed early evidence on these technologies to determine if earlier patient and system access in the NHS is appropriate while further evidence is generated.

The medical technologies advisory committee has considered the evidence and the views of clinical and patient experts. EVA guidance recommendations are conditional while more evidence is generated to address uncertainty in their evidence base. NICE has included advice in this guidance on how to minimise any clinical or system risk of early access to treatment.

Further evidence will be generated over the next 3 years to assess if the benefits of these technologies are realised in practice. NICE guidance will be reviewed to include this evidence and make a recommendation on the routine adoption of these technologies across the NHS.

This document has been prepared for public consultation. It summarises the evidence and views that have been considered and sets out the recommendations made by the committee. NICE invites comments from registered stakeholders, healthcare professionals and the public. This document should be read along with the evidence for this EVA (an EVA report and overview).

The advisory committee is interested in receiving comments on the following:

- Has all of the relevant evidence been taken into account?
- Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?
- Are the recommendations sound, and a suitable basis for guidance to the NHS?

Equality issues

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that the recommendations may need changing to meet these aims. In particular, please tell us if the recommendations:

- could have a different effect on people protected by the equality legislation than on the wider population, for example by making it more difficult in practice for a specific group to access the technologies
- could have any adverse effect on people with a particular disability or disabilities.

Please provide any relevant information or data you have about such effects and how they could be avoided or reduced.

Note that this document is not NICE's final guidance on artificial intelligence technologies to aid contouring for radiotherapy treatment planning. The recommendations in section 1 may change after consultation.

After consultation, NICE will consider the comments received. The final recommendations will be the basis for NICE's early value guidance.

Key dates:

Closing date for comments: **24 August 2023**

1 Recommendations

1.1 Nine artificial intelligence (AI) technologies can be used in the NHS while more evidence is generated to aid contouring for radiotherapy treatment planning in people having external beam radiotherapy. AI technologies must be used with healthcare professional review of contours.

The following technologies can only be used once they have Digital Technology Assessment Criteria (DTAC) approval:

- AI-Rad Companion Organs RT (Siemens Healthineers)
- ART-Plan (TheraPanacea, Oncology Systems)
- DLCExpert (Mirada Medical)
- INTContour (Carina Medical)
- Limbus Contour (Limbus AI, AMG Medtech)
- MIM Contour ProtégéAI (MIM Software)
- MRCAT Prostate plus Auto-contouring (Philips)
- MVision Segmentation Service (MVision AI Oy, Xiel)
- RayStation (RaySearch).

1.2 The technology developers must confirm that agreements are in place to generate the evidence (as outlined in NICE's evidence generation plan) and contact NICE annually to confirm that evidence is being generated and analysed as planned. NICE may withdraw the guidance if these conditions are not met.

1.3 At the end of the evidence generation period (3 years), the technology developers should submit the evidence to NICE in a form that can be used for decision making. NICE will review the evidence and assess if the technologies can be routinely adopted in the NHS.

Evidence generation

1.4 More evidence needs to be generated on the following key outcomes:

- clinical acceptability of contours and amount of edits needed
- impact on radiation dose
- time saving including time for healthcare professional review and edits
- resource use defined by healthcare professional grade and time
- adverse events associated with AI autocontouring and contouring errors.

Potential benefits of early use

- **System benefit:** AI technologies may help healthcare professionals to produce contours more quickly. This may make the workflow more efficient. It may also improve the consistency of contours and increase compliance with national and international guidelines.
- **Clinical benefit:** Clinical evidence suggests that AI technologies generally produce similar quality contours as manual contouring, with most structures needing only minor edits.
- **Resources:** The evidence suggests that AI autocontouring is quicker than manual contouring even when including time for healthcare professional review and edits. This could have potential cost savings. It may also free up healthcare professional time for patient-facing tasks or more complex cases when AI autocontouring may not be appropriate.

Managing the risk of early use

- **Clinical review:** All AI autocontours must be reviewed and edited as needed by a trained healthcare professional before being used in radiotherapy treatment planning.
- **Costs:** Potential cost savings depend on technology costs including setup and maintenance, time saving and the healthcare professional grade of the person doing the contouring. With a band 7 radiographer doing the contouring, cost analysis suggests that the highest priced AI technology (£50 per plan) would need to save around 47 minutes to be cost neutral, compared with 4 minutes for the lowest priced technology (£4 per plan). This guidance will be reviewed within 3 years and the recommendations may change. Take this into account when negotiating the length of contracts and licence costs.
- **Information governance:** Local NHS hospitals and trusts should have appropriate information governance policies for using AI technologies.
- **Equality:** AI models can contain algorithmic bias depending on the population used in training, which may not be representative of populations in clinical

practice. This may affect the performance of AI autocontouring for some populations such as children and young people or people with atypical anatomy.

2 The technologies

2.1 Contouring is an important part of the radiotherapy treatment planning process. It involves outlining the target volumes and organs at risk (OAR) to guide radiotherapy so that treatment is effective and radiation toxicity is minimised. Artificial intelligence (AI) technologies aim to improve contouring efficiency by automatically contouring the OAR and sometimes the target volumes before radiotherapy. They have been trained to process images from CT or MRI scans to produce an initial contour. Images and contours are then reviewed by trained healthcare professionals and modified as needed before use.

2.2 NICE has assessed 11 AI technologies to aid contouring for radiotherapy treatment planning. The criteria for including technologies in this assessment are in the topic scope on the NICE website. Nine technologies have regulatory approval for use in the NHS:

- AI-Rad Companion Organs RT (Siemens Healthineers) is a standalone software that contours over 60 OAR structures on CT scans including the abdomen, head and neck, pelvis and thorax.
- ART-Plan (TheraPanacea, Oncology Systems) is a standalone software that contours over 150 structures including OAR and lymph nodes in the abdomen, brain, head and neck, thorax and pelvis on CT images and the abdomen, brain and male pelvis on MRI scans.
- DLCExpert (Mirada Medical) is deployed on Mirada Medical's Workflow Box platform. It contours over 160 structures on CT and MRI images, including the abdomen, breast, head and neck, prostate and thorax.

- INTContour (Carina Medical) is a standalone software that contours over 60 target and OAR structures from the abdomen, head and neck, male pelvis and thorax.
- Limbus Contour (Limbus AI, AMG Medtech) is a standalone software that contours over 200 OAR and target volumes including lymph nodes, the abdomen, breast, central nervous system, head and neck, lungs, pelvis and prostate on CT images, and the central nervous system, gynaecological and brachy structures on MRI scans.
- MIM Contour ProtégéAI (MIM Software) is a standalone software that contours the head and neck, thorax, lungs and liver, prostate and abdomen structures from CT images and the prostate from MRI scans.
- MRCAT Prostate plus Auto-contouring (Philips) is a clinical application integrated in the Philips Ingenia system for MRI in radiation therapy. It provides automatic contours of the prostate.
- MVision Segmentation Service (MVision AI Oy, Xiel) is a standalone software that contours over 160 structures including OAR and target volumes in the abdomen and thorax, brain, breast, head and neck, and pelvis.
- RayStation (RaySearch) is a radiotherapy external beam and brachytherapy planning system with AI autocontouring functionality. It contours over 70 structures on CT images including the breast and lymph nodes, head and neck, male pelvis, thorax and abdomen.

Two technologies are awaiting CE or UK Conformity Assessed (UKCA) mark approval and so cannot be used yet:

- AutoContour (Radformation) is a standalone software that contours over 200 structures including OAR and lymph nodes in the chest and abdomen, head and neck, and pelvis on CT images and the brain on MRI scans.

- OSAIRIS (Cambridge University Hospitals NHS Foundation Trust) is an open-source standalone software that contours up to 26 head and neck and prostate treatment site structures on CT images.

The Medicines and Healthcare products Regulatory Agency (MHRA) advised that most AI technologies will likely be classified as class IIa or higher under the MHRA guidance on software as a medical device. The government has extended the transition period for CE marked devices in the UK.

Care pathway

2.3 The National Cancer Registration and Analysis Service and Cancer Research UK's statistics on chemotherapy, radiotherapy and surgical tumour resections in England reported that between 2013 and 2019, about 40% of people with cancer had radiotherapy as part of their cancer treatment. Clinical experts advised that AI technologies could be used to aid contouring for everyone having external beam radiotherapy.

2.4 Manual contouring is the most common contouring method in standard care. Manual contouring of target regions is usually done by clinical oncologists. Contouring of OAR may also be done by clinical technologists, dosimetrists or therapeutic radiographers. There are guidelines for contouring OARs and disease sites from organisations such as the European Society for Radiotherapy and Oncology and the Royal College of Radiologists. Other contouring methods include atlas-based contouring and model-based segmentation, but these are not as widely used. Regardless of contouring method, contours should be reviewed before being used in treatment planning in line with guidance such as the Royal College of Radiologists guidance on radiotherapy, target volume definition and peer review.

2.5 AI technologies would be used to provide an initial contour as part of the standard workflow for radiotherapy treatment planning. AI autocontours would then be reviewed by trained healthcare professionals and edited as needed before being used in treatment planning.

The comparators

2.6 Comparators are contouring methods used in standard care to contour OAR and target volumes. These include manual contouring, atlas-based contouring and model-based segmentation. Comparators may also include no contouring for cases when AI technologies may produce contours for structures not routinely contoured in standard care.

3 Committee discussion

NICE's medical technologies advisory committee considered evidence on artificial intelligence (AI) technologies to aid contouring for radiotherapy treatment planning from several sources, including an early value assessment (EVA) report by the external assessment group (EAG), and an overview of that report. Full details are in the project documents for this guidance.

Unmet need and potential value

3.1 AI autocontouring with healthcare professional review may be quicker than other contouring methods, which could reduce healthcare professional time to do contouring. This could reduce costs and increase efficiency, which may increase capacity, allow more focus on patient-facing tasks and reduce waiting lists. The Royal College of Radiologists clinical oncology census report 2021 reported workforce pressure because of staff shortages and continued effects from the COVID-19 pandemic. Clinical experts advised that they spend a lot of time creating and reviewing manual contours. They said that healthcare professionals have reported finding it easier to review and edit AI autocontours than to create contours from scratch.

3.2 Clinical experts advised that AI technologies could improve the consistency of contours and compliance with national and international guidelines. Some AI technologies have been trained using guidelines and may be regularly updated when guidelines update. One expert said that AI technologies helped improve how they were defining structures. AI technologies may also produce smoother contours of 3D structures compared with manual contouring.

Clinical effectiveness

Potential benefits

3.3 The committee considered that there was strong evidence for the potential usefulness of AI technologies to aid contouring in radiotherapy treatment planning. The relevant evidence consisted of 79 studies, including 27 full-text publications and 52 conference abstracts. Because of the large number of publications, the EAG extracted data from 15 prioritised studies, specifically:

- 8 prospective studies (DLCExpert, Limbus Contour, MIM Contour ProtégéAI and MRCAT Prostate plus Auto-contouring)
- 5 retrospective studies, (AI-Rad Companion Organs RT, INTContour, MVision Segmentation Service, OSAIRIS, RayStation)
- 2 conference abstracts (ART-Plan, AutoContour).

The level of evidence varied across technologies, but all technologies had some evidence showing potential benefits of AI autocontouring. Overall, the clinical evidence showed that AI autocontours were generally similar to manual contours, with most rated as clinically acceptable and ready to use or needing only minor edits. AI autocontouring was also consistently quicker than manual contouring even when including time for healthcare professional review and edits. The committee concluded that AI autocontouring with healthcare professional review and edits was likely to be clinically equivalent to manual contouring and quicker to do.

Differences in performance

3.4 The evidence showed that while AI autocontouring worked well for most organs at risk (OAR) and clinical target volumes, there were some structures that needed major edits or were unusable. These were typically smaller structures such as the cochlea, optic chiasm, optic nerve, penile bulb and pituitary gland. The clinical experts advised that AI autocontouring performed similarly in clinical practice. AI technologies sometimes have difficulties contouring very small or irregularly shaped organs. AI autocontours may also be less accurate in cases with atypical anatomy or trouble with positioning during imaging. One expert estimated that for head and neck structures, about 90% to 95% of AI autocontours would be accurate. The clinical experts noted that over time, healthcare professionals learn where specific AI technologies produce less accurate contours. This means they can make edits more quickly because they know that certain areas of the contour are likely to need editing.

Cost and resource use

3.5 Cost consequences analysis showed that potential cost saving from using AI autocontouring as an alternative to manual contouring depended on technology costs, time saving and healthcare professional grade of the person doing the contouring:

- Technology costs ranged from £4 to £50 per plan and included software (licence and subscription), hardware, data storage, and upgrade and maintenance costs. Several companies advised that healthcare professional training is also included in these costs.
- The clinical evidence reported time savings ranging from 3 minutes to 80 minutes, but the EAG advised that these savings did not always include the time for healthcare professional review and edits. The clinical experts estimated time saving of 10 minutes to 30 minutes depending on the amount of editing needed. The committee noted the

importance of clinical acceptability of the AI autocontours because this may affect the number of edits needed when reviewed by a healthcare professional.

- Experts advised that contouring of OAR may be done by band 6 or 7 radiographers or registrars if there are not enough radiographers. Contours are usually reviewed by consultant clinical oncologists. But there may be many people involved in the review and sign-off of contours, which may make it difficult to estimate true resource use.

3.6 The simple cost offset calculator showed that as technology costs increased, the time saving needed for AI technologies to be cost saving or cost neutral also increased. The same was found for healthcare professional grade needed to do manual contouring. For example:

- With the lowest technology cost of £4 per plan and a band 7 radiographer (£65 per hour based on PSSRU 2021) doing the contouring, time saved would need to be around 4 minutes for the AI technology to be cost neutral.
- With the highest technology cost of £50 per plan and a band 7 radiographer doing the contouring, the time saved must be around 47 minutes for the AI technology to be considered cost neutral.

The EAG advised that there were several factors in the analysis that may cause a wide variation in results. It noted that the limited cost-effectiveness evidence made it difficult to draw firm conclusions about the potential cost effectiveness of the time saving compared with manual contouring. Estimates of healthcare professional costs may also vary depending on the source used. The committee concluded that while there were uncertainties in the cost analysis, AI technologies were likely to be cost saving or cost neutral but this largely depended on the technology costs and time saving.

Implementation

Managing risk

3.7 AI technologies would be used to aid contouring for radiotherapy treatment planning within the existing care pathway. AI autocontours must always be reviewed by trained healthcare professionals and edited as needed before being used. There were no adverse events reported in the evidence or by the clinical experts. So, the committee considered that the risk of AI autocontouring with healthcare professional review and edits is likely to be low. Some companies said they have tools that healthcare professionals can use to report errors in the performance and outputs of their technology. The committee concluded that there should be ongoing reporting of any errors in AI autocontouring and adverse events associated with these technologies.

3.8 Local NHS hospitals and trusts should have appropriate information governance policies for using AI technologies. AI technologies should also have national and local Digital Technology Assessment Criteria (DTAC) approval before being used in clinical practice.

3.9 In the future, more widespread use of these technologies could result in a skill loss in the workforce. Clinical experts advised that this was unlikely because healthcare professionals would nearly always do some editing as part of their review of the autocontours. Also, some technologies provided training packages for healthcare professionals to practice their skills.

Technical considerations

3.10 The committee considered that the compatibility of AI technologies with current systems may vary in each NHS hospital or trust. Most technologies were identified as being available to work with any system and so should have minimal technical implementation issues. The experts advised that AI technologies should be compatible with existing hospital systems if they use the DICOM (Digital Imaging and Communications in Medicine) format.

Equality considerations

3.11 AI models can contain algorithmic bias depending on the population used in training, which may not be representative of populations in clinical practice. This may cause bias based on age, disability, sex and geographical location. Experts advised that there may be a lack of representation of female pelvis and breast cancer in men in some training datasets. Training datasets may also underrepresent children and young people.

3.12 AI technologies used to aid contouring may work best with certain CT or MRI sequences or with the person being in a specific position. Training datasets may not include data on atypical positioning or atypical anatomy, for example, if someone has had a previous surgery. This may affect how well AI autocontouring works for these populations. Healthcare professionals may consider manual contouring to be more appropriate for some people because it may produce more accurate contours in these specific cases. This is not thought to affect patient care or outcomes but may affect time to produce, review and edit contours.

3.13 Risk of bias should be considered as part of a local assessment process when deciding if to use AI technologies. Companies should provide information on training datasets as part of their product information pack. Clinical experts advised that most AI technologies were not retrained on local training sets although some had the capacity for this. Ideally in the future AI technologies could be trained on a representative national population.

Evidence gap overview

3.14 For all technologies, evidence gaps can be related to the population, the intervention, or the main outcomes. The committee concluded that there was enough evidence of potential benefits from the 11 technologies for them to be used in the NHS once they have appropriate regulatory approvals, while further evidence is generated to address these gaps. Important evidence gaps for these technologies are:

- Population: the most assessed anatomical sites were the head and neck, and pelvis or prostate. More evidence is needed on how well AI autocontouring works in different anatomical sites. There was no relevant published evidence on using AI autocontouring in specific population groups, such as children and young people or people with atypical anatomy because of surgery. The committee considered that the demographics of datasets used for training an algorithm may differ from populations in clinical settings. It highlighted the need for evidence generation in how AI technologies work in clinical practice in local populations to ensure no specific group is disadvantaged.
- Intervention: there were 2 technologies (ART-Plan and AutoContour) with no full-text evidence.
- Outcomes: only 4 technologies (DLCExpert, INTContour, Limbus Contour and RayStation) had evidence that included dosimetric analysis. The committee highlighted the need for further evidence on dosimetric analysis in other technologies. In all technologies there was a need for long-term patient outcomes and data on adverse events including the impact of AI autocontouring on radiation toxicity. There were 6 technologies (AI-Rad Companion Organs RT, DLCExpert, Limbus Contour, MIM Contour ProtégéAI, MVision Segmentation Service, OSAIRIS) that had evidence on the time saved using AI technologies compared with atlas-based or manual contouring. Time saved was highlighted as a key potential benefit of these technologies.

The committee highlighted the importance for evidence generation in timesaving to include time for healthcare professional review and edits after AI technology use. Clinician acceptability and number of edits needed to AI contour may impact overall time saved and so should be accounted for. The committee also noted that the time and cost saving potential is impacted by who edits the contour. So, any evidence generated should include healthcare professional grade and the impact of this on time and cost saving.

3.15 In addition to the key outcomes listed in section 1.2, the committee agreed that real-world evidence on using AI autocontouring in clinical practice could provide more valuable information about:

- accuracy and acceptability of autocontours across a range of anatomical sites
- how well AI technologies work in an NHS population, including people with limited mobility or atypical anatomy
- the frequency of software updates and impact of updates on how well AI autocontouring works.

In the longer term, evidence on patient outcomes such as radiation toxicity could become available.

4 Committee members and NICE project team

Committee members

This topic was considered by NICE's medical technologies advisory committee, which is a standing advisory committee of NICE.

Committee members are asked to declare any interests in the technologies to be evaluated. If it is considered there is a conflict of interest, the member is excluded from participating further in that evaluation.

The minutes of the medical technologies advisory committee meetings, which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Additional specialist committee members took part in the discussions and provided expert advice for this topic:

Specialist committee members

Dr Angela Pathmanathan

Consultant clinical oncologist, Royal Marsden NHS Foundation Trust

Dr Carl Rowbottom

Consultant clinical scientist and director of physics, The Clatterbridge Cancer Centre NHS Foundation Trust

Dr David Bernstein

Lead physicist in radiotherapy imaging, Royal Marsden NHS Foundation Trust

Dr Jacqueline Poxon

Deputy head of treatment planning physics, Barts Health NHS Trust

Mr Mayur Munshi

Lead for dosimetry, Beatson West of Scotland Cancer Centre

Dr Samantha Warren

Consultant clinical scientist, Newcastle upon Tyne University Hospitals NHS Trust

Mr Syed Moinuddin

Lead research and development radiographer, UCLH

Dr Theresa Urbano

Consultant clinical oncologist and adjunct reader, Guy's and St Thomas's NHS Foundation Trust and King's College London

NICE project team

Each medical technologies guidance topic is assigned to a team consisting of 1 or more health technology assessment analysts (who act as technical leads for the topic), a health technology assessment adviser and a project manager.

Dionne Bowie and Alice Bell

Health technology assessment analyst and associate health technology assessment analyst

Bernice Dillon

Health technology assessment adviser

Elizabeth Islam

Project manager

ISBN: