NICE National Institute for Health and Care Excellence

Evidence generation plan for artificial intelligence technologies to aid contouring for radiotherapy treatment planning

Implementation support

Published: 27 September 2023

www.nice.org.uk

Contents

1 Purpose of this document 3

2 Evidence gaps 4

2.1 Essential evidence for future committee decision making 4

2.2 Evidence that further supports committee decision making 5

3 Approach to evidence generation 6

3.1 Ongoing studies 6

3.2 Data sources 8

3.3 Evidence collection plan 9

3.4 Data to be collected 9

3.5 Evidence generation period 11

4 Monitoring 12

5 Implementation considerations 13

1 Purpose of this document

NICE's assessment of artificial intelligence technologies to aid contouring for radiotherapy treatment planning recommends that further evidence is generated while they are being used in the NHS.

This plan outlines the evidence gaps and what real-world data needs to be collected for a NICE review of the technologies again in the future. It is not a study protocol.

The technology developers are responsible for ensuring that data collection and analysis takes place. Support for evidence generation will be available through a competitive process facilitated by the Office for Life Sciences, pending business case approval. This will be in the form of funding for evidence generation consortia, bringing analytical partners and implementation sites together with developers for evidence generation.

Guidance on commissioning and procurement of the technology will be provided by NHS England, who are developing a digital health technology policy framework to further outline commissioning pathways.

NICE will withdraw the guidance if the technology developers do not meet the conditions about monitoring evidence generation in section 4 on monitoring.

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

2 Evidence gaps

This section describes the evidence gaps, why they need to be addressed and their relative importance for future committee decision making.

The committee will not be able to make a positive recommendation without the essential evidence gaps (see section 2.1 on essential evidence for future committee decision making) being addressed. The companies can strengthen the evidence base by also addressing as many other evidence gaps (see section 2.2 on evidence that further supports committee decision making) as possible. Addressing these other evidence gaps will help the committee to make a recommendation by ensuring it has a better understanding of the patient or healthcare system benefits of the technology.

2.1 Essential evidence for future committee decision making

Time saving and resource use

To estimate the time-saving benefits of the technologies, it is important to measure the total time needed for contouring. This is because a reduction in the time taken to review radiotherapy contours could reduce patient waiting times by, allowing healthcare professionals to review more patients in a day. In addition, the evidence generated should capture the perceived impact on time saving and other factors that may influence this.

To understand how the software affects resource use, evidence should be generated on the time spent by healthcare professionals on reviewing and editing the software's contours compared with manual or atlas-based contouring. Given that there is substantial variation in time saving between different anatomical structures, information should be collected and presented for each anatomical structure.

Further information is needed on the NHS pay bands of the reviewing healthcare professionals to inform the cost calculations. To estimate total cost, it is also important to understand the cost of training, implementing the software and related administration.

Organ delineation and acceptability of the contour

Acceptability of the software's outputs can be measured using a Likert-type scale considering the number of edits by the reviewing healthcare professional. Fewer or only minor edits would suggest greater acceptability. To complement this, information about the settings included in the software and the guidelines they align with should be included.

To further understand how the technology provides contours that are clinically acceptable, it is essential to capture the experiences and opinions of healthcare professionals during data collection.

2.2 Evidence that further supports committee decision making

Adverse effects of treatment

Artificial intelligence (AI) contouring could improve patient outcomes by more accurately delineating organs, leading to more accurate treatment and fewer adverse effects of treatment. Information on adverse effects of treatment and dosimetric analyses comparing the AI technologies with manual or atlas-based contouring should be collected.

Performance in different anatomical sites and patient subgroups

To better understand the benefits of AI contouring compared with manual or atlas-based contouring, generating evidence on the technologies' performance on anatomical sites other than head, neck and prostate is advised.

Also, evidence should be generated about the software's performance when contours may be challenging to obtain because, for example, a person has limited mobility or atypical anatomy. Subgroup analysis for these people can be done by collecting patient-level information on age, sex, ethnicity, relevant comorbidities or disabilities and the anatomical sites targeted by the scan.

This analysis is important if the developer expects their technology to be used for other anatomical areas and patient subgroups.

3 Approach to evidence generation

An approach to generating evidence for artificial intelligence (AI) contouring is presented. How this will address the evidence gaps is considered, and any strengths and weaknesses highlighted.

Most technologies do not have ongoing studies that will address the evidence gaps. The King's Technology Evaluation Centre are doing a study (completing by June 2024) with some of the technologies, which may address:

- organ delineation and acceptability of the contour
- time saving and resource use
- performance in different anatomical sites.

3.1 Ongoing studies

Table 1 summarises the evidence gaps and ongoing studies that might address them.

Table 1 Summary of the evidence gaps and ongoing studies

Evidence gaps and technologies	Time saving and resource use	Organ delineation and acceptability of the contour	Adverse effects of treatment	Performance in different anatomical sites and patient subgroups
AI-Rad Companion Organs RT (Siemens Healthineers)	Evidence is available Ongoing study	Evidence is available Ongoing study	No relevant evidence identified	No relevant evidence identified

Evidence gaps and technologies	Time saving and resource use	Organ delineation and acceptability of the contour	Adverse effects of treatment	Performance in different anatomical sites and patient subgroups
ART Plan (Thera-Panacea)	No relevant evidence identified Ongoing study	No relevant evidence identified Ongoing study	No relevant evidence identified	No relevant evidence identified Ongoing study
DLCExpert (Mirada Medical)	Evidence is available	Evidence is available	No relevant evidence identified	Limited available evidence
INTContour (Carina Medical)	No relevant evidence identified	Evidence is available	No relevant evidence identified	No relevant evidence identified
Limbus Contour (Limbus AI, AMG Medtech)	Evidence is available Ongoing study	Evidence is available Ongoing study	No relevant evidence identified	Evidence is available Ongoing study
MIM Contour Protege AI (MIM Software)	Evidence is available	Evidence is available	No relevant evidence identified	No relevant evidence identified
MRCAT Prostate plus Auto-contouring (Philips)	No relevant evidence identified	No relevant evidence identified	No relevant evidence identified	No relevant evidence identified

Evidence gaps and technologies	Time saving and resource use	Organ delineation and acceptability of the contour	Adverse effects of treatment	Performance in different anatomical sites and patient subgroups
Mvision Segmentation Service (Mvision AI Oy, Xiel)	Evidence is available Ongoing study	Evidence is available Ongoing study	No relevant evidence identified	Evidence is available Ongoing study
RayStation (RaySearch)	Evidence is available Ongoing study	Evidence is available Ongoing study	No relevant evidence identified	Evidence is available Ongoing study

Information about current evidence status is derived from the external assessment group's report; evidence not meeting the scope and inclusion criteria are not included. AutoContour (Radformation) does not currently have any evidence for the evidence gaps.

3.2 Data sources

There are several data collections that have different strengths and weaknesses that could potentially support evidence generation. NICE's real-world evidence framework provides detailed guidance on assessing the suitability of a real-world data source to answer a specific research question.

The Radiotherapy Data Set (RTDS) is the national standard for collecting radiotherapy data in the NHS. It is currently collecting data for all NHS Acute Trust providers of radiotherapy services in England. It will need to be modified to collect data addressing the evidence gaps. But this could take up to 2 years.

Local or regional data collections such as the sub-national secure data environments that measure outcomes specified in the evidence generation plan could be used to collect data to address the evidence gaps. Secure data environments are data storage and access platforms that bring together many sources of data, such as from primary and secondary

care, to enable research and analysis. The sub-national secure data environments are designed to be agile and can be modified to suit the needs of new projects.

The quality and coverage of real-world data collections are of key importance when used in generating evidence. Active monitoring and follow-up through a central coordinating point is an effective and viable approach of ensuring good-quality data with high coverage.

3.3 Evidence collection plan

To address the evidence gaps, a before and after study is suggested. A before and after design allows comparison when there are differences between sites or departments, such as in processes, protocols or equipment. In a before and after study, data is collected and compared before and after implementing the AI contouring technologies in radiotherapy departments.

Data collection for a particular technology can be at a single centre or ideally across multiple centres.

For qualitative outcomes, surveys or interviews could be used to assess people's experiences and views on the technologies' acceptability, performance and impact on productivity in routine clinical practice. Open-ended questions could be included to gather information on the potential of the technologies to improve current clinical practice.

3.4 Data to be collected

Outcome variables for data collection should include data for the technology and, where appropriate, the current standard of care (manual or atlas-based contouring). The following outcomes have been identified for collection through the suggested before and after studies:

Quantitative

Information to be collected before and after implementation:

- Total time needed for contouring.

- Average number of contours completed per hour per reviewer.
- NHS band of the reviewer.
- Characteristics of patients reviewed. For example, age, sex, ethnicity, height and weight or body mass index, and comorbidities that may make scans challenging to perform.
- Adverse events and dosimetric analyses.

Information to be collected after implementation:

- Acceptability of the contours, measured by a scale:
  - score 0: no edits needed
  - score 1: minor edits needed
  - score 2: moderate edits needed
  - score 3: major edits needed.
- Training, implementation, and administrative costs.

Qualitative

- Perceived impact on time to review and edit contours.
- Information about other factors that may influence time saving in clinical practice.
- Perceived ease of use.
- Perceived acceptability of output and accuracy of contours.
- Variability of contour accuracy in groups for whom contours may be more challenging to do.
- How use of the technology may affect contouring skills.
- Opinion on patient outcomes:
  - accurate organ delineation
  - accurate clinical target volume delineation

- improvements in throughput
- adverse events.

Information about the technologies

Information about how the technologies were developed and the effect of updates should also be collected. See the NICE evidence standards framework for guidance.

3.5 Evidence generation period

This will be 3 years to allow for setting up, implementation, data collection, analysis, and reporting.

4 Monitoring

Technology developers are required to contact NICE:

- within 6 months of the evidence generation plan's publication to confirm that agreements are in place to generate the evidence specified
- annually to confirm that the data is being collected and analysed as planned.

Technology developers should inform NICE at the earliest opportunity of anything that may affect ongoing evidence generation, including:

- any substantial risk that the evidence will not be collected as planned
- new safety concerns
- significant changes to the technology that affect the evidence generation process.

If data collection is expected to end later than planned, the technology developers should contact NICE to arrange an extension to the evidence generation period. NICE reserves the right to withdraw the guidance if data collection is delayed, or if it is unlikely to resolve the evidence gaps.

5 Implementation considerations

Developers should work with providers and central NHS England teams to begin evidence generation. Planning for a prespecified period for the set-up of the technology is advised. During this period, training and implementation should be done before data collection is started, to account for learning effects. The following considerations around implementing the evidence generation process have been identified through working with system partners:

- Technology developers should provide training for staff in using the artificial intelligence (AI) software.
- Sites should be carefully selected to, where appropriate, maximise data collection for subgroups of interest.

The following barriers for implementing the evidence generation process have been identified through working with system partners:

- the availability of research funds for data collection, analysis and reporting
- the availability of NHS funding to cover the costs of implementing the technology in clinical practice
- lack of expertise and staff to collect data
- burden on clinical staff; the need to have training before implementation, data collection and follow-up
- differences in practice between large tertiary referral centres and smaller hospitals
- variation in treatment protocols and equipment between centres
- the software may be incompatible with other computer packages and scanners used in the NHS
- the availability and ability of NHS information technology departments to install the software.

ISBN: 978-1-4731-5416-2