


Beyond boundaries: exploring radiographers' experiences and solutions in organ-at-risk delineation for radiotherapy planning

Antoine Attard and Susan Mercieca 

Department of Radiography, Faculty of Health Sciences, University of Malta, Msida, Malta

Original Article

Cite this article: Attard A and Mercieca S. (2024) Beyond boundaries: exploring radiographers' experiences and solutions in organ-at-risk delineation for radiotherapy planning. *Journal of Radiotherapy in Practice*. 23(e3), 1–8. doi: [10.1017/S1460396923000432](https://doi.org/10.1017/S1460396923000432)

Received: 25 July 2023

Revised: 22 November 2023

Accepted: 6 December 2023

Keywords:

Auto-contouring; interobserver variation; manual contouring; organs at risk

Corresponding author:

Susan Mercieca;

Email: susan.mercieca@um.edu.mt

Abstract

Introduction: This study aimed to assess the perception of radiographers on the difficulties encountered during the delineation of organs at risk (OARs) for radiotherapy planning and the methods that could be used to facilitate this process.

Methodology: A self-designed questionnaire was distributed to all radiographers ($n = 29$) employed at the radiotherapy department in Malta. The survey assessed the challenges faced by the radiographers during the delineation of various OARs and inquired about the perceived effectiveness of specific measures that could be used to facilitate the delineation process using a Likert scale ranging from 1 (not difficult/effective) to 5 (very difficult/effective).

Results: The response rate was 79%. Overall, the delineation of OARs was rated as slightly difficult (mean score 1.95 ± 0.33). Nevertheless, some OARs, such as the parotid, stomach and brain stem, were deemed more difficult to contour with a mean score of 3 or higher. The implementation of peer review was perceived as the most influential factor in reducing delineation difficulty (mean score 4.59 ± 0.58), followed by contouring training (mean score 4.48 ± 0.58) and training on artificial intelligence (AI) (mean score 4.35 ± 0.48).

Conclusion: The introduction of peer review, training and AI could facilitate the delineation of OARs.

Background

The delineation of the gross tumour volume (GTV) and organs at risk (OARs) on computed tomography (CT) images is an important part of the radiotherapy planning process. Failure to accurately define the GTV and OARs could result in a geographical tumour miss and unnecessary dose to normal tissue, eventually leading to poor local control and more side effects. Despite the importance of this process, the manual definition of the GTV and OARs is prone to observer variation and has often been described as 'the weakest link in radiotherapy'^{1,2}. Various factors have been found to contribute to observer variation. The contouring difficulty depends on the structure being contoured. For example, some structures, like the lung and bladder, have clear boundaries that allow for consistent delineation, while others, such as the parotids and small bowel, present ambiguous borders, leading to higher variability in interpretation among different observers³. Moreover, the healthcare professionals responsible for contouring may have different training backgrounds^{4,5}, experience levels⁶ and visual perception skills, leading to variations in identifying and outlining specific structures. In many radiotherapy centres, the delineation of OARs is often delegated to radiographers or dosimetrists. However, the clinical oncologist is ultimately responsible for checking and approving the contours^{2,4,5}. Although this division of roles can streamline the planning process, it can further contribute to the variability in contouring practices across different institutions.

Studies have shown that protocols⁷, training⁸ and peer review² can reduce the contouring variability of both tumours and OARs. Several automated tools have been developed to facilitate the contouring process. Atlas-based autosegmentation (ABAS) deforms atlases of anatomy previously defined on a reference image onto a new patient image to create a new structure set that fits the patient's anatomy. Numerous studies have found that this method can reduce the delineation time and the interobserver variability in the delineation of some OARs^{9,10}. Nevertheless, the main limitation of this technology is that the department needs to develop a library of plans to be used for the contouring process¹⁰. However, recently, the introduction of artificial intelligence (AI), in particular the development of deep learning technology, is showing great potential in addressing the challenges associated with manual contouring. Numerous deep learning-based commercial contouring solutions with pre-trained models are now available on the market¹¹. Studies have shown that deep learning technology can further improve the accuracy of the contours when compared with ABAS^{12,13}. However, despite the potential of AI technology in improving consistency and reducing the contouring time, the implementation of this technology in clinical practice has been slow mainly due to challenges

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



encountered during the training and validation of these algorithms, variable accuracy results, lack of resources, staff acceptance and data protection issues^{11,14}.

In our department, the radiographers perform the contouring of most OARs, while straightforward GTVs are contoured by higher specialist trainee doctors, and complex GTVs are delineated by the clinical oncologist. For complex cases, a radiologist is sometimes consulted if deemed necessary by the clinical oncologist. However, the clinical oncologist is ultimately responsible for checking and approving all the contours before treatment planning. A recent internal audit revealed significant levels of interobserver variation (average Hausdorff distance of 11.7mm) in the definition of the GTV for patients undergoing head and neck radiotherapy. This has triggered the need to perform a consultation process to evaluate what can be done to improve the contouring process at our local department. Therefore, this study aimed to evaluate the challenges encountered by radiographers working at the local oncology department during the contouring of specific OARs. In addition, the study also evaluated the perception of radiographers on techniques such as AI, training and peer review that could be used to address this problem.

Methodology

Research design

The study used a quantitative, prospective, non-experimental, cross-sectional research design with a self-administered survey.

Development of the data collection tool

The online questionnaire was developed by the researcher using the Google Forms platform. The questionnaire content was informed by relevant literature findings^{2,3} and was divided into five parts (Supplementary Data 1). The initial section consisted of six questions designed to assess the professional background, contouring experience and training received by the radiographers. In the second section, the radiographers were asked to rate the difficulties encountered during the contouring of specific OARs. In addition, the radiographers were also asked to rate their agreement with specific statements related to contouring challenges (e.g., delineation time, workflows and subjectivity). The third and fourth sections assessed the current contouring tools used by the radiographers and their perception of the use of AI software in delineating OARs. In the last and final section, the radiographers were asked to indicate their agreement with the introduction of measures (e.g., time allotted to the task, training, peer review and the introduction of AI software) that could be used to facilitate the contouring process. In addition, an open-ended question was included to allow participants to offer further insights and suggestions regarding methods that could be implemented to facilitate and improve the quality of OARs delineation at the local radiotherapy department. Sections A, C and the last part of section D consisted of multiple-choice questions, while sections B, the first part of section D and section E consisted of Likert scale-type questions whereby the radiographers were asked to rate the difficulty levels or agreement to specific statements using a 5-point Likert scale ranging from 1 (not difficult/do not agree) to 5 (very difficult/strongly agree).

Validation of the questionnaire

Four academic radiographers (two Maltese, one from Ireland and one from England) with more than 5 years of experience were

asked to evaluate every question of the questionnaire for clarity and relevance in relation to the aims and objectives of the study using a Likert scale ranging from 1 (not clear/not relevant) to 4 (very clear/highly relevant). Furthermore, the lecturers were asked to make suggestions to improve the questionnaire. The content validity index for each question (I-CVI) was then used to determine the level of agreement between the four lecturers. The I-CVI was calculated by dividing the number of items rated as relevant (3) or highly relevant (4) by the total number of experts. The average content validity of the questionnaire was calculated by dividing the total I-CVI by the total number of questions. The CVI score ranges from 0 to 1, and a score of 0.78 or higher indicates a good level of agreement¹⁵. All items within the questionnaire had an I-CVI of 0.8 or higher, and the average CVI for the instrument was 0.88, which indicates that the questionnaire is relevant. The overall clarity score for the questionnaire was 0.83. However, in response to the expert comments, certain questions were reworded to improve clarity.

Reliability testing and pilot study

Prior to the data collection process, two radiographers were asked by The Society of Radiographers Malta (SRM) to test the questionnaire for intra-rater reliability. The radiographers were asked to complete the same questionnaire twice within a 1-week interval. Given that the sample size for the study was predicted to be small, the same radiographers who tested for questionnaires for reliability were also used as part of the pilot study to ensure the smoothness of the data collection process.

Kendall's tau test was used to assess intra-rater reliability. The Kendall's tau test resulted in a coefficient (τ) of 0.98, with a p-value less than $p < 0.001$, indicating satisfactory test-retest reliability. The radiographers did not comment on the questionnaire or data collection process; thus, no changes were made to the questionnaire¹⁶.

Data collection

The target population for this study consisted of all radiographers (therapy or dually qualified) working at the radiotherapy department in Malta. The accessible population included all members of the target population who were willing to complete the online questionnaire. The data collection was conducted between February 2023 and March 2023. The SRM acted as an intermediary for the study and invited all radiographers working in Malta to participate in this study via email. The email explained the purpose of the study and provided the radiographers with a weblink to complete the questionnaire. No incentive was offered to complete the questionnaire. However, multiple reminders were distributed by the SRM to improve the response rate.

Ethical considerations

The Faculty of Health Science Research Ethics Committee (FREC) of the University of Malta granted approval to perform this study. No personally identifiable information was collected in the questionnaire, and completing the questionnaire constituted consent to participate in the study.

Data analysis

The Statistical Package for Social Sciences software was used to analyse the data. The categorical data were summarised as percentages, while the continuous data were summarised as means \pm standard deviations. The Kruskal Wallis test was used to evaluate the impact of experience

Table 1. The professional background of the participants

	N	%
Total number of participants	23	100%
Experience		
Above 5 years	15	65.22%
Below 5 years	8	34.78%
Performance of contouring duties		
Every day	3	13.04%
Almost every day	4	17.39%
Occasionally	12	52.17%
Rarely	4	17.39%
Additional contouring training		
Yes	3	13.04%
No	20	86.96%

(less than 5 years, more than 5 years) and frequency of practising contouring (rarely, occasionally, almost every day and every day) on the OARS mean difficulty score. For all statistical tests, a p-value below 0.05 was deemed statistically significant. The open-ended question was evaluated using content analysis.

Results

The professional background of the participants

The local radiotherapy department currently employs 29 radiographers, of whom 23 (79.3%) completed the questionnaire. The professional characteristics of the participants are summarised in Table 1. Most of the participants (65.22%) had more than 5 years of experience working as radiographers at the local oncology hospital. However, most participants (52.17%) declared that they were occasionally assigned contouring duties. Only three (13.04%) of the participants attended additional post-graduate training on contouring. These courses were provided by professional bodies (n = 2) or were conducted as part of a Master of Science in Radiotherapy (n = 2).

Perceived difficulty level contouring specific OARs

The mean difficulty rating score for all OARs was 1.95 ± 0.33 , which indicates that the radiographers identified slight difficulties when contouring the OARs. However, some OARs, such as the parotid, stomach and brain stem, were found to be somewhat difficult to contour, with a mean rating score of 3 or higher. The mean OARs difficulty delineation score for each organ is shown in Fig. 1.

Factors affecting the perceived delineation difficulty for specific OARs

Table 2 summarises the impact of experience on the perceived difficulty level contouring specific OARs. Overall, the radiographers with more than 5 years of experience found the contouring of OARs slightly less difficult than radiographers with less than 5 years of experience (mean score: 2.03 versus 1.87 $p = 0.241$). However, the difference was not statistically significant. The same pattern was observed for the majority (14/20) of the OARs.

Table 3 summarises the impact of duty time allotted to contouring on the perceived difficulty level of contouring specific OARs. The radiographers who were frequently (mean score: 1.82 ± 0.3) assigned contouring duties found the contouring of OARs slightly less difficult than radiographers who were assigned contouring duties occasionally (mean score: 1.88 ± 0.21) or rarely (mean score: 2.38 ± 0.35). However, the difference was marginally statistically significant ($p = 0.055$).

Evaluation of factors that lead to contouring challenges

The time required to contour OARs was perceived as the most challenging factor (mean score 4.52 ± 0.65), followed by distraction with other duties (mean score 4.38 ± 0.70), long screen time (mean score 4.00 ± 0.64) and subjectivity of contouring (mean score 3.90 ± 0.55). Overall, the radiographers agreed that the time allotted by management to perform the task was sufficient (mean score 3.90 ± 0.62). With regard to the mean rating score related to the sufficiency of experience, training and image quality and the suitability of the contouring software, the mean rating score was shifted towards the neutral side with a mean score ranging from 3.14 to 3.44 (Fig. 2).

Tools used by the radiographers to contour OARs

All radiographers made use of manual and semi-automated tools to contour the OARs. None of the radiographers used AI-based software to contour OARs. However, the majority (n = 18, 78.3%) of the radiographers reported being aware of the availability of AI software that could be used to contour OARs, and only 5 (21.74%) reported having used the Elekta ABAS AI software, which is available at the department. However, it is interesting to note that the ABAS software does not use an AI algorithm to contour OARs but rather a deformable image registration software that deforms the anatomy from a user-defined reference atlas onto a new patient's CT data. Two of the radiographers rated this software as good, and the rest rated it as average.

Radiographers' perception of the use of AI software for contouring OARs

Although this technology was not used locally, most radiographers (n = 18, 78.26%) were aware of the existence of AI software for contouring OARs. The perception of radiographers on the use of AI software is summarised in Fig. 3. Overall, the radiographers agreed that the AI software was too laborious to train. However, the radiographers disagreed that AI software is generally very accurate and that the contours generated required minimal editing. The radiographers did not find data protection an issue for the use of AI software. Moreover, they indicated that they lacked the training and confidence to use such software.

Techniques that could be used to facilitate the contouring process

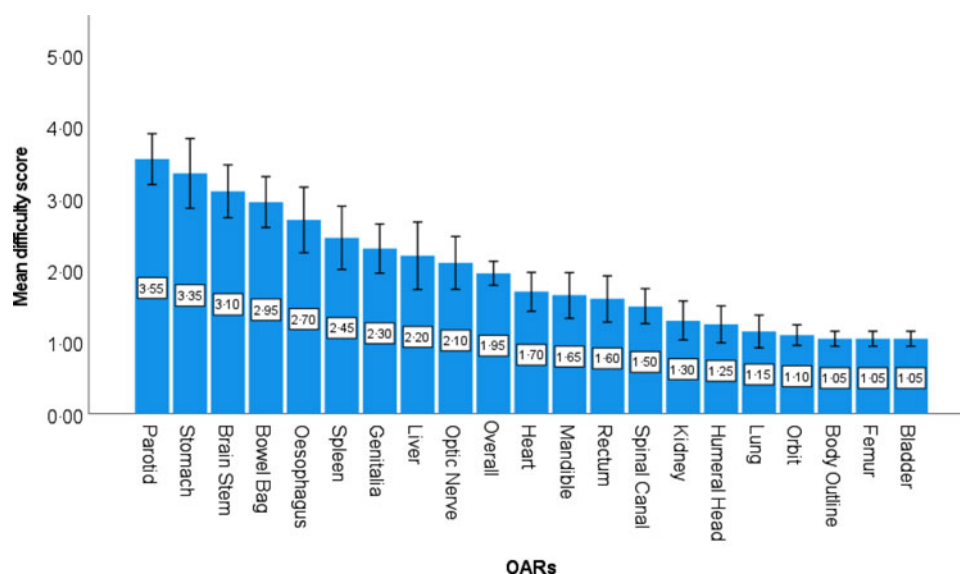
The radiographers agreed that the introduction of peer review, training on contouring, training on AI, internal and external audits, more detailed protocols and the purchasing of automated contouring software could facilitate and improve the contouring of OARs with all statements receiving a mean score of 3.9 or higher (Fig. 4).

A total of seven participants provided additional suggestions to improve the contouring of OARs at the local oncology department. One participant emphasised the necessity of adhering to national

Table 2. Impact of experience on the participants' perceived difficulty contouring specific OARs

OAR	Experience	N	Mean score	Std	p-Value	OAR	Experience	N	Mean score	Std	p-Value
Genitalia	<5 years	11	2.09	0.70	0.147	Kidney	>5 years	11	1.45	0.69	0.104
	>5 years	12	2.73				<5 years	12	1.08		
Bladder	<5 years	11	1.09	0.30	0.296	Lung	>5 years	11	1.27	0.65	0.131
	>5 years	12	1.00				<5 years	12	1.00		
Bowel Bag	<5 years	11	3.00	0.77	1	Liver	>5 years	11	2.45	1.13	0.081
	>5 years	12	3.00				<5 years	12	1.67		
Heart	<5 years	11	1.73	0.65	0.620	Spleen	>5 years	11	2.82	0.87	0.103
	>5 years	12	1.58				<5 years	12	2.18		
Femur	<5 years	11	1.09	0.30	0.296	Stomach	>5 years	11	3.36	1.03	0.974
	>5 years	12	1.00				<5 years	12	3.25		
Body outline	>5 years	11	1.09	0.30	0.296	Brain stem	>5 years	11	3.00	0.89	1.000
	<5 years	12	1.00				<5 years	12	3.00		
Spinal canal	>5 years	11	1.45	0.52	0.677	Mandible	>5 years	11	1.64	0.67	0.499
	<5 years	12	1.55				<5 years	12	1.83		
Oesophagus	>5 years	11	2.82	1.08	0.729	Parotid	>5 years	11	3.73	0.79	0.218
	<5 years	12	2.64				<5 years	12	3.36		
Optic nerve	>5 years	11	2.00	0.89	0.461	Humeral	>5 years	11	1.40	0.70	0.185
	<5 years	12	2.27				<5 years	12	1.08		
rectum	>5 years	11	1.73	0.79	0.209	Orbit	>5 years	11	1.36	0.92	0.852
	<5 years	12	1.33				<5 years	12	1.17		
Total	>5 years	11	2.03	0.43	0.241						
	<5 years	12	1.87								

OAR, organs at risk; n, number of participants; Std, standard deviation. A score of 1 indicates not difficult and a score of 5 indicates very difficult.

**Figure 1.** The mean difficulty rating score for each OARs and overall. A value of 1 indicates not difficult, while a value of 5 indicates very difficult.

contouring guidelines. In addition, three respondents emphasised the importance of additional training on contouring and AI. Four of the participants highlighted the need to introduce AI contouring

software, with one of the participants stressing that this is now becoming increasingly important with the imminent implementation of the MR-Linac within the local radiotherapy department.

Table 3. Impact of duty time assigned to contouring on the perceived difficulty level contouring specific OARs

OARs	Time-assigned contouring duties	N	Mean Score	Std	p-Value	OARs	Time-assigned contouring duties	N	Mean Score	Std	p-Value	
Genitalia	Rarely	4	2.25	0.43	0.129	Kidney	Rarely	4	1.75	0.83	0.146	
	Occasionally	12	2.75	1.01			Occasionally	12	1.17	0.37		
	Frequently/daily	7	1.83	0.37			Frequently/daily	7	1.14	0.35		
Bladder	Rarely	4	1.25	0.43	0.059	Lung	Rarely	4	1.50	0.87	0.188	
	Occasionally	12	1.00	0.00			Occasionally	12	1.00	0.00		
	Frequently/daily	7	1.00	0.00			Frequently/daily	7	1.14	0.35		
Bowel bag	Rarely	4	3.50	0.50	0.184	Liver	Rarely	4	2.75	1.09	0.093	
	Occasionally	12	3.08	0.64			Occasionally	12	1.83	0.99		
	Frequently/daily	7	2.57	0.73			Frequently/daily	7	2.00	0.76		
Heart	Rarely	4	1.75	0.43	0.374	Spleen	Rarely	4	3.00	0.71	0.599	
	Occasionally	12	1.58	0.49			Occasionally	12	2.42	0.76		
	Frequently/daily	7	1.71	0.70			Frequently/daily	7	2.33	1.11		
Femur	Rarely	4	1.25	0.43	0.059	Stomach	Rarely	4	3.75	1.09	0.103	
	Occasionally	12	1.00	0.00			Occasionally	12	3.08	1.04		
	Frequently/daily	7	1.00	0.00			Frequently/daily	7	3.43	0.73		
Body Outline	Rarely	4	1.25	0.43	0.059	Brain Stem	Rarely	4	3.50	0.87	0.086	
	Occasionally	12	1.00	0.00			Occasionally	12	2.83	0.69		
	Frequently/daily	7	1.00	0.00			Frequently/daily	7	3.00	0.76		
Spinal Canal	Rarely	4	1.75	0.43	0.060	Mandible	Rarely	4	2.25	0.43	0.074	
	Occasionally	12	1.33	0.47			Occasionally	12	1.83	0.69		
	Frequently/daily	7	1.67	0.47			Frequently/daily	7	1.29	0.45		
Oesophagus	Rarely	4	3.25	1.09	0.262	Parotid	Rarely	4	4.00	0.71	0.553	
	Occasionally	12	2.67	0.85			Occasionally	12	3.45	0.66		
	Frequently/daily	7	2.50	0.96			Frequently/daily	7	3.43	0.73		
Optic nerve	Rarely	4	3.00	0.00	0.081	Humeral Head	Rarely	4	1.33	0.47	0.865	
	Occasionally	12	2.00	0.71			Occasionally	12	1.167	0.373		
	Frequently/daily	7	1.83	0.69			Frequently/daily	7	1.29	0.70		
Rectum	Rarely	4	2.00	0.71	0.116	Orbit	Rarely	4	2.25	1.09	0.003*	
	Occasionally	12	1.5	0.65			Occasionally	12	1.000	0.000		
	Frequently/daily	7	1.29	0.45			Frequently/daily	7	1.14	0.35		
Total	Rarely	4	2.38	0.35	0.055							
	Occasionally	12	1.88	0.21								
	Frequently/daily	7	1.82	0.30								

OAR, organs at risk; n, number of participants; Std, standard deviation. A score of 1 indicates not difficult and a score of 5 indicates very difficult. * indicates statistically significant.

Figure 2. Agreement of the radiographers towards statements related to challenges encountered while contouring OARs. A score of 1 indicates a strong disagreement with the statement, while a score of 5 indicates a strong agreement.

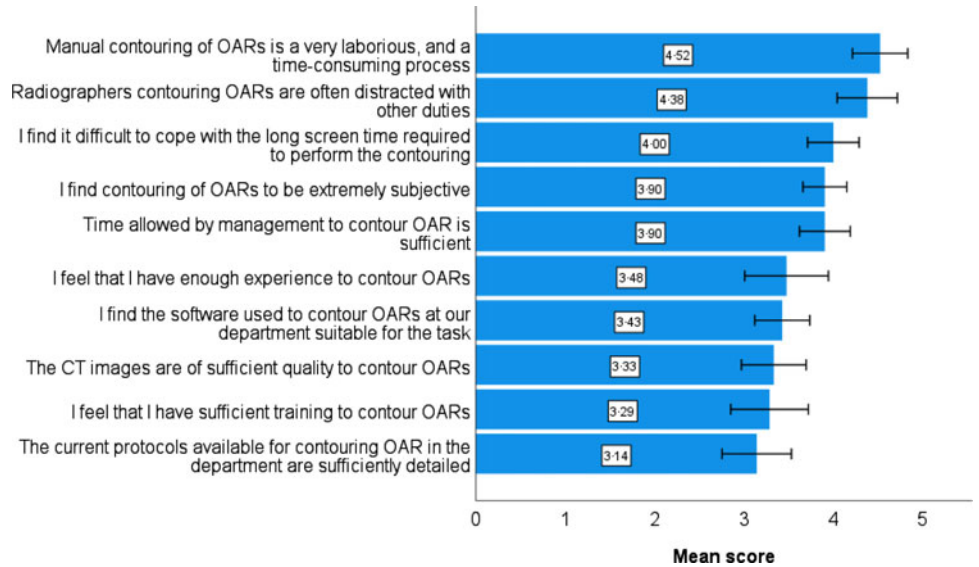


Figure 3. Perception of radiographers on the use of AI software for the contouring of OARs. A score of 1 indicates a strong disagreement with the statement provided, while a score of 5 indicates a strong agreement.

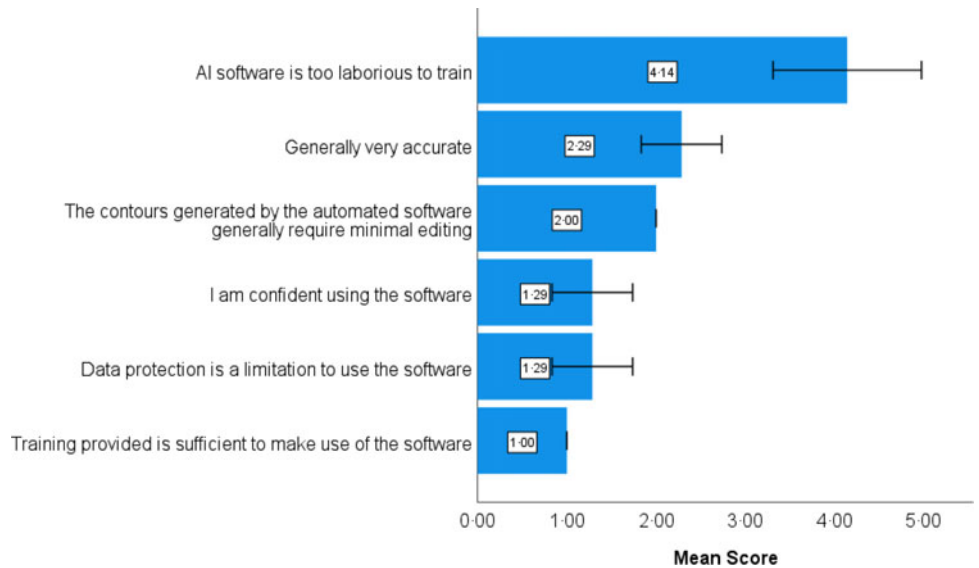
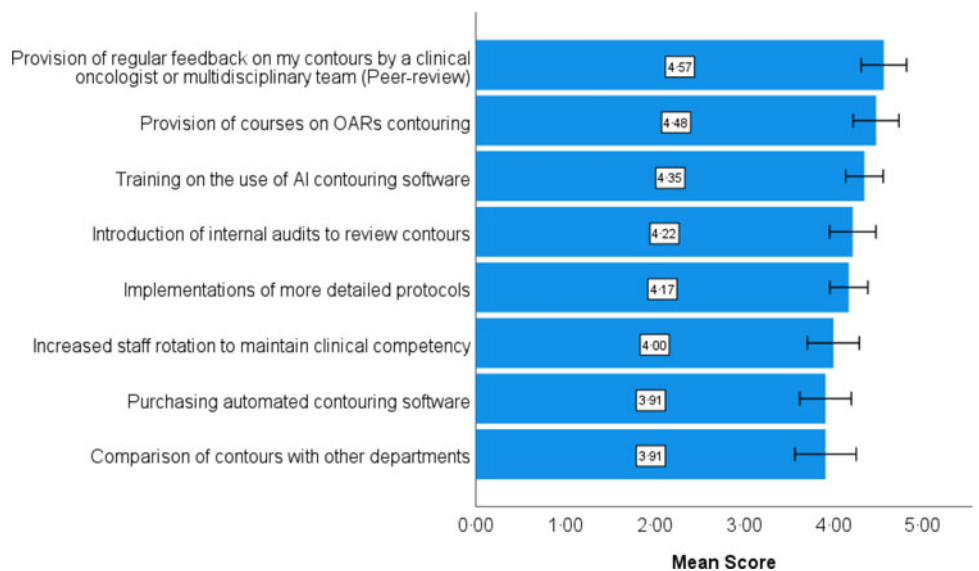


Figure 4. Radiographers' perception on the introduction of techniques that could be used to facilitate and improve the delineation of OARs. A score of 1 indicates a strong disagreement with the statement provided, while a score of 5 indicates a strong agreement.



Discussion

The challenges radiographers encounter during the definition of OARs can vary between departments. Since the delineation of OARs is an emerging role for radiographers, very few studies have been published on the matter^{17,18}. In this study, we performed the first survey to evaluate the contouring challenges encountered by radiographers at the local oncology hospital in Malta. The findings from our survey hold significant implications for informing the hospital management about the necessity of implementing novel workflows, training programmes and best practices in OAR delineation, particularly with the imminent introduction of the MR-Linac.

As expected, the radiographers found more difficulties contouring OARs with more complex shapes and boundaries, such as the parotid, small bowel, stomach and brain stem, when compared with structures with less complex shapes and boundaries, such as the femur, bladder and body outline. Studies have found that the level of experience⁶ and professional background of clinical oncologists¹⁹ can impact the consistency of the contours. In our studies, radiographers with less than 5 years of experience and those who were rarely assigned delineation of OARs duties found the contouring task slightly more difficult. However, the difference was not statistically significant for both parameters, possibly due to the small sample size.

All radiographers received most of their OAR contouring training during their undergraduate course programme. Only three radiographers pursued further post-graduate training, either through a Master's degree programme or by enrolling in a course provided by a professional body. Due to the limited number of participants, it was not possible to statistically evaluate the impact of training on the challenges encountered contouring OARs. It is important to acknowledge that multiple training interventions may be required to reach an adequate level of competency in contouring²⁰. Moreover, despite the wide range of contouring courses available, most contouring training courses have been predominantly targeted towards oncologists, and there is a notable scarcity of courses specifically designed to enhance radiographers' contouring skills. As a result, radiographers may have limited opportunities to receive specialised training in this critical aspect of radiation therapy planning¹⁷.

The radiographers found the introduction of peer review as the most needed intervention to improve the quality of the OARs delineation in the department. While multiple professional bodies advocate for the implementation of peer review to reduce the occurrence of gross contouring errors and facilitate training interventions to address contouring inconsistencies, not all departments possess adequate resources to incorporate this process into routine clinical practice^{2,21,22}.

The long contouring time and the long screen time were identified as the main challenges when contouring OARs. More frequent staff rotations could enable radiographers to maintain their skills while minimising the fatigue caused by prolonged screen time. Studies have shown that the introduction of AI-based contouring may also reduce the contouring time and reduce the observer variation^{12,13}. Although most radiographers were aware of the availability of automated software to contour OARs, only five radiographers reported using such software. However, when asked to elaborate, the radiographers stated that they used ABAS, which does not use AI to delineate OARs. A licence for ABAS has been available in the department for many years. Nevertheless, this software was never implemented clinically as the vendors did not

provide atlases with the system, and the radiographers lacked the expertise and found the software too laborious to train. Overall, the radiographers did not have highly positive views on using AI for contouring OARs. The radiographers agreed that AI models are very laborious to train and often generate contours that lack accuracy and require too much editing. This negative perception might have been based on the department's experience using ABAS. Recent studies have shown that deep learning algorithms can reduce the delineation time and improve the accuracy of OARs contours compared to manual contouring¹³. Moreover, AI software vendors are now providing trained AI models that can be more easily implemented within the radiotherapy workflows¹¹.

The findings of our survey carry significant implications for the clinical practices at our local oncology department. Survey responses underscore the imperative need to establish an ongoing training programme within the department aimed at ensuring that all radiographers maintain their competency in this specialised field. To effectively design and implement this programme, collaboration among oncologists, physicists and radiographers with expertise in this domain is essential. The proposed training course could be structured as a self-study programme, wherein radiographers are tasked with contouring OARs across a variety of clinical cases. In order to enhance learning, automatic feedback mechanisms should be integrated, utilising quantitative metrics to provide radiographers with real-time performance evaluations. Radiographers who do not meet the specified criteria for proficiency in contouring should receive additional training and guidance from senior radiographers or clinical oncologists. In addition, regular contouring audits should be conducted as part of the quality assurance programme of the department. These audits would enable the management to identify specific areas that require improvement within the department's contouring practices. Furthermore, whenever feasible, OAR contouring should be a focal point of discussion during multidisciplinary team meetings. Finally, as AI-based contouring is expected to become more established, there is clearly a need for the hospital to invest in this technology and provide training to enable radiographers to use this technology safely.

This study has some limitations that have to be acknowledged. The small sample size in our study may have limited the statistical significance of our research findings, particularly for the subgroup analysis. Moreover, we did not gauge the perception of dosimetrists and doctors on the contouring challenges encountered while contouring OARs. Therefore, larger multicentre studies that include all professionals involved in the contouring process are required to improve the generalisability of the research findings. Although some open-ended questions were included, the survey was predominately based on closed-ended questions, which may have limited the collection of in-depth information about the subject. In order to avoid having a very long questionnaire, some of the questions aimed at gauging a general overview of the contouring challenge for all OARs rather than asking detailed questions for each OAR. As a result, we could not gain a full understanding of the specific challenges encountered by the radiographers for the delineation of each of the OARs. Finally, surveys rely on the participants' perceptions and self-assessment of their contouring practices. Without direct observation, it is not possible to validate contouring accuracy or identify discrepancies between self-reported practices and actual behaviour. It is important to note that the majority of the questionnaires were completed by radiographers who were not regularly involved in delineation and lacked previous exposure to ABAS or other deep-

learning contouring systems. As a result, some participants may have had difficulty recalling the challenges encountered during the contouring process. Moreover, they may have provided an inaccurate rating for the system based on perception rather than their actual experience or true knowledge. Therefore, further research is recommended to assess the interobserver variation within the department and to assess the impact of introducing peer review and training on the quality of the OARs contours. Further studies should make use of interviews to gain more in-depth information about the subject. Moreover, workshops could be performed to enable the radiographers to gain a better understanding of this technology. Further studies should be performed to assess the true impact of these workshops on the radiographers' confidence and contouring accuracy should be assessed.

Conclusion

Overall, the radiographers encountered slight difficulties when contouring OARs. Nevertheless, some OARs, such as the parotid, stomach and brain stem, were deemed more difficult to contour. The introduction of peer review and training were identified as the two key factors required to improve the quality of the contours. The introduction of AI-based contouring software could also reduce the delineation time. However, training is essential to facilitate its smooth introduction into clinical workflows.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S1460396923000432>.

Author contributions. Mr. Antoine Attard collected the data, analysed the data and wrote the manuscript. Dr. Susan Mercieca supervised the project and edited the final manuscript.

Competing interests. The authors have no conflict of interest to declare.

References

- Njeh CF. Tumor delineation: the weakest link in the search for accuracy in radiotherapy. *J Med Phys* 2008; 33: 136–40. <https://doi.org/10.4103/0971-6203.44472>.
- The Royal College of Radiologists. Radiotherapy target volume definition and peer review. RCR Guidance. 2022: 1–35. <https://www.rcr.ac.uk/publication/radiotherapy-target-volume-definition-and-peer-review> (accessed May 15, 2018).
- Mercieca S, Belderbos JSA, Herk M. Challenges in the target volume definition of lung cancer radiotherapy. *Transl Lung Cancer Res* 2020; 10: 1983–199. <https://doi.org/10.21037/TLCR-20-627>.
- Arculeo S, Miglietta E, Nava F, et al. The emerging role of radiation therapists in the contouring of organs at risk in radiotherapy: analysis of inter-observer variability with radiation oncologists for the chest and upper abdomen. *Ecancermedicallscience* 2020;14: 996. <https://doi.org/10.3332/ECANCER.2020.996>.
- Shessel A, Rosewall T. Quantifying the geometric and dosimetric impact of inter-observer variability for radiation oncologists and radiation therapists when contouring the organs at risk within the upper abdomen. *J Med Imaging Radiat Sci* 2022; 53: S7. <https://doi.org/10.1016/J.JMIR.2022.04.022>.
- Giraud P, Elles S, Helfre S, et al. Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. *Radiother Oncol* 2002; 62: 27–36. [https://doi.org/10.1016/S0167-8140\(01\)00444-3](https://doi.org/10.1016/S0167-8140(01)00444-3).
- Lin D, Lapen K, Sherer MV, et al. A systematic review of contouring guidelines in radiation oncology: analysis of frequency, methodology, and delivery of consensus recommendations HHS public access. *Int J Radiat Oncol Biol Phys* 2020; 107: 827–35. <https://doi.org/10.1016/j.ijrobp.2020.04.011>.
- Tsang Y, Hoskin P, Spezi E, et al. Assessment of contour variability in target volumes and organs at risk in lung cancer radiotherapy. *Tech Innov Patient Support Radiat Oncol* 2019; 10: 8–12. <https://doi.org/10.1016/J.TIPRSO.2019.05.001>.
- Tao CJ, Yi JL, Chen NY, et al. Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in nasopharyngeal carcinoma: a multi-institution clinical study. *Radiother Oncol* 2015; 115: 407–11. <https://doi.org/10.1016/J.RADONC.2015.05.012>.
- Hu Y, Byrne M, Archibald-Heeren B, et al. Implementing user-defined atlas-based auto-segmentation for a large multicentre organisation: the Australian Experience. *J Med Radiat Sci* 2019; 66: 238–49. <https://doi.org/10.1002/jmrs.359>.
- Robert C, Munoz A, Moreau D, et al. Clinical implementation of deep-learning based auto-contouring tools – experience of three French radiotherapy centers. *Cancer/Radiothérapie* 2021; 25: 607–16. <https://doi.org/10.1016/J.CANRAD.2021.06.023>.
- Lustberg T, van Soest J, Gooding M, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiother Oncol* 2018; 126: 312–7. <https://doi.org/10.1016/J.RADONC.2017.11.012>.
- van Dijk L V., Van den Bosch L, Aljabar P, et al. Improving automatic delineation for head and neck organs at risk by Deep Learning Contouring. *Radiother Oncol* 2020; 142: 115–23. <https://doi.org/10.1016/j.radonc.2019.09.022>.
- Victor Mugabe K. Barriers and facilitators to the adoption of artificial intelligence in radiation oncology: a New Zealand study. *Tech Innov Patient Support Radiat Oncol* 2021; 18: 16–21. <https://doi.org/10.1016/J.TIPRSO.2021.03.004>.
- Gilbert GE, Prion S. Making sense of methods and measurement: Lawshe's Content Validity Index. *Clin Simul Nurs* 2016; 12: 530–1. <https://doi.org/10.1016/j.ecns.2016.08.002>.
- Laerd Statistics. The ultimate IBM® SPSS® Statistics guides. Lund Lund Res Ltd 2018. <https://statistics.laerd.com/> (accessed February 13, 2020).
- Oliveira C, Barbosa B, Couto JG, et al. Advanced practice roles amongst therapeutic radiographers/radiation therapists: a European survey. *Radiography* 2023; 29: 261–73. <https://doi.org/10.1016/J.RADI.2022.12.003>.
- Oliveira C, Barbosa B, Couto JG, Bravo I, Khine R, McNair H. Advanced practice roles of therapeutic radiographers/radiation therapists: a systematic literature review. *Radiography* 2022; 28: 605–19. <https://doi.org/10.1016/J.RADI.2022.04.009>.
- Peng Y, Chen L, Shen G, et al. Interobserver variations in the delineation of target volumes and organs at risk and their impact on dose distribution in intensity-modulated radiation therapy for nasopharyngeal carcinoma. *Oral Oncol* 2018; 82: 1–7. <https://doi.org/10.1016/J.ORALONCOLOGY.2018.04.025>.
- Konert T, Vogel W V., Everitt S, et al. Multiple training interventions significantly improve reproducibility of PET/CT-based lung cancer radiotherapy target volume delineation using an IAEA study protocol. *Radiother Oncol* 2016; 121: 39–45. <https://doi.org/10.1016/j.radonc.2016.09.002>.
- Chera BS, Mazur L, Adams RD, Marks LB. The promise and burden of peer review in radiation oncology. *J Oncol Pract* 2016; 12: 196–8. <https://doi.org/10.1200/JOP.2015.008961>.
- Caissie A, Rouette J, Jugpal P, et al. A pan-Canadian survey of peer review practices in radiation oncology. *Pract Radiat Oncol* 2016; 6: 342–51. <https://doi.org/10.1016/J.PPRO.2016.01.014>.