Physics and Imaging in Radiation Oncology 34 (2025) 100765

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Original Research Article

Multicentre prospective risk analysis of a fully automated radiotherapy workflow

Geert De Kerf [a,b,*], Ana Barragán-Montero [c], Charlotte L. Brouwer [d], Pietro Pisciotta [d], Marie-Claude Biston [e], Marco Fusella [f], Geoffroy Herbin [g], Esther Kneepkens [h], Livia Marrazzo [i], Joshua Mason [j], Camila Panduro Nielsen [k,l], Koen Snijders [a], Stephanie Tanadini-Lang [m], Aude Vaandering [n], Tomas M. Janssen [o]

[a] Department of Radiation Oncology, Iridium Netwerk, Wilrijk (Antwerp), Belgium
[b] Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium
[c] UCLouvain – Institut de Recherche Expérimentale et Clinique - Molecular Imaging Radiotherapy and Oncology (MIRO), Brussels, Belgium
[d] University of Groningen, University Medical Center Groningen, Department of Radiation Oncology, Groningen, the Netherlands
[e] Centre Léon Bérard, 28 rue Laennec 69373 LYON Cedex 08, France
[f] Abano Terme Hospital, Department of Radiation Oncology, Abano Terme (Padua), Italy
[g] Ion Beam Applications, Louvain-la-Neuve, Belgium
[h] Department of Radiation Oncology (Maastro, GROW School for Oncology, Maastricht University Medical Center+, Maastricht, the Netherlands
[i] University of Florence, Department of Experimental and Clinical Biomedical Sciences, Florence, Italy
[j] Imperial College Healthcare NHS Trust, London, UK
[k] Laboratory of Radiation Physics, Department of Oncology, Odense University Hospital, Odense, Denmark
[l] Department of Clinical Research, University of Southern Denmark, Odense, Denmark
[m] Department of Radiation Oncology, University Hospital Zurich and University of Zurich, Zurich, Switzerland
[n] Radiation Therapy Department, Cliniques Universitaires St Luc, Brussels, Belgium
[o] Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Automation
FMEA
Auto-contouring
Auto-planning
Computer assisted radiotherapy planning

ABSTRACT

Background and Purpose: Fully automated workflows (FAWs) for radiotherapy treatment preparation are feasible, but remain underutilized in clinical settings. A multicentre prospective risk analysis was conducted to support centres in managing FAW-related risks and to identify workflow steps needing improvement.
Material and Methods: Eight European radiotherapy centres performed a failure mode and effect analysis (FMEA) on a hypothetical FAW, with a manual review step at the end. Centres assessed occurrence, severity and detectability of provided, or newly added, failure modes to obtain a risk score. Quantitative analysis was performed on curated data, while qualitative analysis summarized free text comments.
Results: Manual review and auto-segmentation were identified as the highest-risk steps and the highest scoring failure modes were associated with inadequate manual review (high detectability and severity score), incorrect (i.e. outside of intended use) application of the FAW (high severity score) and protocol violations during patient preparation (high occurrence score). The qualitative analysis highlighted amongst others the risk of deviation from protocol and the difficulty for manual review to recognize automation errors. The risk associated with the technical parts of the workflow was considered low.
Conclusions: The FMEA analysis highlighted that points where people interact with the FAW were considered higher risk than lack of trust in the FAW itself. Major concerns were the ability of people to correctly judge output in case of low generalizability and increasing skill degradation. Consequently, educational programs and interpretative tools are essential prerequisites for widespread clinical application of FAWs.

* Corresponding author at: Department of Radiation Oncology, Iridium Netwerk, Oosterveldlaan 24, Antwerp, Belgium.
E-mail address: geert.dekerf@zas.be (G. De Kerf).

https://doi.org/10.1016/j.phro.2025.100765
Received 22 January 2025; Received in revised form 2 April 2025; Accepted 3 April 2025
Available online 6 April 2025

1. Introduction

Cancer incidence is increasing across Europe, accompanied by rising healthcare costs and a shrinking (medical) workforce [1]. To ensure that every patient can receive the highest quality of care within acceptable waiting times, automation of radiotherapy treatment preparation is often regarded as essential [2]. Recent years have seen advancements in artificial intelligence (AI) applications within the field of radiotherapy. Tools have been developed capable of image segmentation, dose optimization, outcome prediction, quality assurance, image reconstruction, etc. These innovations enable the automation of individual tasks, offering potential efficiency gain [3–5]. However, in clinical practice, manual checks are still performed after each automated task, limiting the potential benefit [6,7]. Nonetheless, automation has proven its power to reduce risks if adequate training was part of the implementation [8–12].

Larger efficiency gains can be achieved when the entire workflow is automated. A fully automated workflow (FAW) would eliminate intermediate manual checks, handovers and waiting times between tasks. Technically, the tools required to implement a FAW for radiotherapy treatment preparation are currently available [13]. Yet, despite the adoption of task-specific automation, such as segmentation and treatment planning, limited studies report on the practicalities of implementing FAWs and they have not been widely adopted in clinical practice [14–16]. Nevertheless, similar concerns arise within the online adaptive setting, where safely implementing automation is even more stringent [17,18].

Suggested barriers that hinder the implementation of FAWs have been documented in literature, such as a lack of appropriate trust in the automation tools, uncertainties regarding the human role in FAWs, unclear responsibilities for quality control and clinical decisions, and the complexity of implementing innovations in workflows in clinical practice [19–21]. To overcome these barriers, the clinical introduction of FAWs requires the development of new safeguards for monitoring and quality assurance of both input and output [22,23]. Additionally, defining the optimal workflow involves balancing the roles of medical professionals, processes, and technology [24].

The development of these safeguards and workflows should be motivated by a suitable analysis of the risks associated with the FAW. To address this challenge, a multi-centre, prospective risk analysis for the fully automated radiotherapy workflow was initiated to support centres in managing FAW-related risks, identify workflow steps requiring additional tools or improvements and provide insights to facilitate the clinical adoption of FAWs. To this end, a Failure Mode and Effect Analysis (FMEA) was performed for a hypothetical FAW [25]. This analysis aims to identify potential key barriers to implementation, paving the way for broader adoption of FAWs in clinical practice.

2. Materials and methods

2.1. Multicentre setup

An open call for participation was made during the ESTRO physics workshop 2023 and eight European radiotherapy centres and one company responded to that call and were included in the project: Iridium Netwerk (Antwerp, Belgium), Cliniques Universitaires Saint Luc (Brussels, Belgium), University Medical Center Groningen (Groningen, The Netherlands), Abano Terme Hospital (Abano Terme, Italy), Careggi University Hospital (Florence, Italy), Maastro (Maastricht, The Netherlands), Odense University Hospital (Odense, Denmark),The Netherlands Cancer Institute (Amsterdam, The Netherlands), and IBA (Ion Beam Applications, Louvain-la-Neuve, Belgium). The company was present during the group discussions and this participation is in line with the ESTRO physics workshop policy to invite company representatives as equal peers.

To gauge the differences in experience in automation between centres, a questionnaire was sent out and all eight centres used auto-segmentation of organs at risk, while only five centres did report any experience on target auto segmentation. Seven centres reported experience with auto-planning, but workflow automation was only applied by five centres as summarized in Supplementary Material 1.

2.2. Hypothetical workflow

Before initiating the FMEA, a hypothetical FAW for radiotherapy treatment preparation was developed over multiple sessions (Fig. 1). The hypothetical workflow concerned an offline treatment preparation (e.g. not online/adaptive) and was intentionally tumour site-agnostic.

The workflow assumed that all treatment preparation was done using CT imaging only, with no additional imaging and image registration required. Furthermore, it was assumed that following patient referral by the radiation oncologist, the patient was assigned to the automated workflow by selecting an FAW treatment. From that point, aside from CT simulation, all steps were assumed fully automated and handled by the FAW orchestrator. Finally, at the end of the automation, a manual review was performed by both a radiation oncologist and medical physicist. Apart from the manual review at the end, the FAW explicitly includes no automated QA or sanity checks.

A detailed description of the hypothetical workflow, its underlying assumptions, and a high-level description of the presumed software architecture was shared with the participating centres and is provided in the Supplementary Material 2.

2.3. FMEA

Based on the description of the hypothetical workflow, an FMEA template was made, outlining each step and sub-step of the FAW and its potential failures. The FMEA was designed following the specifications of the AAPM TG-100 report [26].

During a breakout session at the 2023 ESTRO physics workshop, causes (what is causing the failure), potential failure modes (the way something fails), and effects (consequence of the failure) were identified. Combined with the analysis of Nealon et al. [9], known failure modes and effects were provided in the FMEA template, to prevent duplication. The template was reviewed and approved by all participants and is provided in the Supplementary Material 3.

The severity (S, impact of the failure), occurrence (O, likelihood of the failure) and detectability (D, likelihood of the failure reaching the patient undetected) were left open. The template was shared with all centres who were asked to form a multi-disciplinary team to score the parameters for each failure mode. The multi-disciplinary team had to consist of key users, which depended on the local organization of the department. Participants were encouraged to identify additional failure modes and to provide comments where necessary.

Each multi-disciplinary team conducted the FAW in the context of their clinical routine, as summarized in Supplementary Material 1, and consisted of at least 1 radiation oncologist, 1 RTT and 2 medical physicists. Local FMEA's were organized in a two months time period (May till June 2024).

All filled-in FMEA templates are provided as Supplementary Material 4.

2.4. Quantitative analysis

Data curation was performed on the collected FMEA data to facilitate multicentre comparisons. Based on group discussions, potential causes and effects were grouped into five and eight major categories, respectively. The potential causes included:

1. Human error (e.g. unexpected use of contrast, patient identification error)

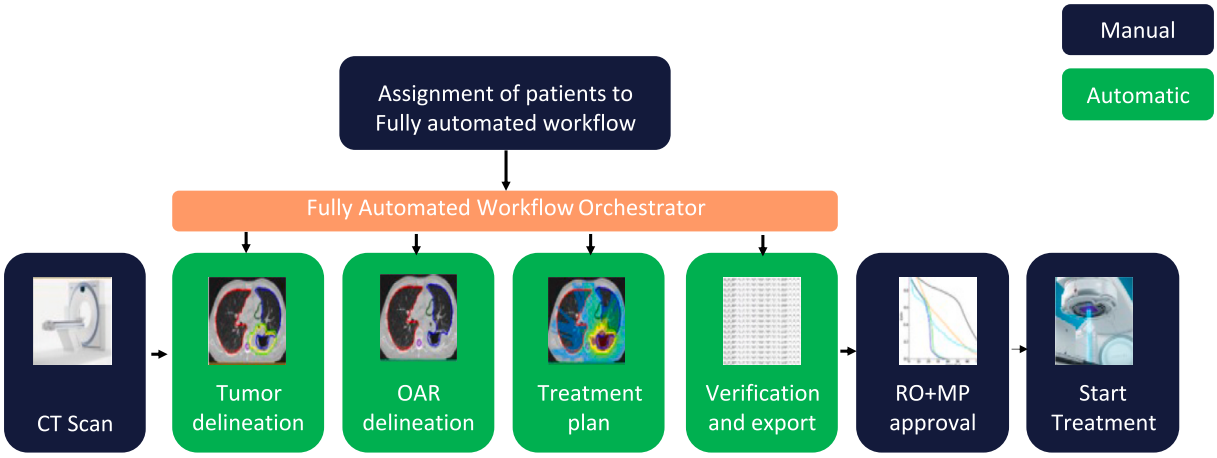

Fig. 1. Schematic overview of the FAW with automatic and manual workflow steps respectively marked in green and blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2. Lack of generalizability of AI-based algorithms (e.g. out of distribution bladder volume)
3. Patient specific factors (e.g. patient implant, unexpected use of contrast) inadequate bladder filling)
4. Scan details (e.g. incorrect information in CT dicom tag)
5. Software errors (e.g. failing automation triggers, incorrect placement of isocentre)

The potential effects were grouped as follows:

1. FAW stalls
2. Inadequate quality auto-segmentation
3. Incorrect patient setup vector (e.g. wrong detection of CT localization point)
4. Incorrect plan documentation (e.g. wrong information in pdf plan report)
5. Suboptimal treatment plan created
6. Incorrect treatment delivered
7. Communication gap (e.g. incorrect documentation of FAW assignement)
8. Treatment delays

Next, to minimize variation in scoring, for every centre the scoring of O, S and D was normalized to the respective maximum value of that centre, obtaining O*, S* and D*. The risk R* was recalculated based on the normalized values ($R^* = O^* \times S^* \times D^*$), reducing the impact of the institution-dependent risk assessment related to differences in equipment, level of standardization and training,… Since centres were allowed to add new risks to the FMEA, single, highly scored risks could disproportionately influence the analysis. To address this, only failure modes scored by at least two centres were included in the quantitative analysis.

To identify the process step with the highest risk in an automated workflow, the maximum R* per process step was determined for each centre and the average over the centres was compared between workflow steps.

The failure modes, causes, and effects were analysed by identifying the risks with the highest R*, averaged over all centres. To unravel how R* was composed, the relative contribution of each individual, average scoring of O*, S* and D* was plotted next to its corresponding R* value. The failure modes analysis was limited to the five highest scoring modes for each of the metrics (R*, O*, S* and D*).

2.5. Qualitative analysis

In addition to the quantitative analysis, all comments provided by the participants were collected and analysed. The comments were structured and grouped according to the main workflow steps of the hypothetical FAW. The summary of the re-grouped comments was subsequently reviewed and approved by all participants.

3. Results

Each centre's normalized maximum risk per process step showed substantial variation in scoring risks across centres (Fig. 2). The highest identified risks were observed in the manual review and auto-segmentation step with an average maximum R* of $\mu = 0.46$ (standard deviation, $\sigma = 0.17$) and $\mu = 0.44$ ($\sigma = 0.20$) over the different centres respectively. Risks related to auto-planning ($\mu = 0.25$, $\sigma = 0.18$), simulation ($\mu = 0.25$, $\sigma = 0.13$), and patient assignment ($\mu = 0.18$, $\sigma = 0.14$) followed. The pitfalls related to the FAW orchestrator were assessed as minimal risk ($\mu = 0.07$, $\sigma = 0.07$).

The top five highest-scoring values for O*, S*, D* and R*, along with the associated failure mode, revealed that the highest identified risks were related to inadequate manual review of the automated plan and auto-segmentation, driven by high D* and S* (Fig. 3). Treating the wrong patient within an automated workflow had an exceedingly high S*, but both O* and D* were perceived as low. Another high S* was given to the incorrect prescription assignment, leading to an incorrect run of the FAW. Suboptimal protocol assignment during the FAW, resulting in clinically inadequate segmentation or plan quality, carried high O* and S* scores, but were perceived as detectable (low D*).

As a potential cause, the generalizability of AI models emerged as causing the highest risk. On the contrary, the software itself did not hold a substantial risk as shown in Fig. 4.

The potential effects are summarized in Fig. 5, and most important effects were a suboptimal auto-segmentation and treatment plan creation. Effects with the highest estimated O* were communication gaps and time delays. An incorrect treatment delivery got the highest S* score.

The qualitative review (Table 1) highlighted skill degradation concerns in both auto-segmentation and auto-planning. Within the manual review step, the risk of a failing manual error detection was seen as a significant challenge in FAWs, while within the patient assignment step the major concern related to applications of the FAW outside its intended, highly standardized use.

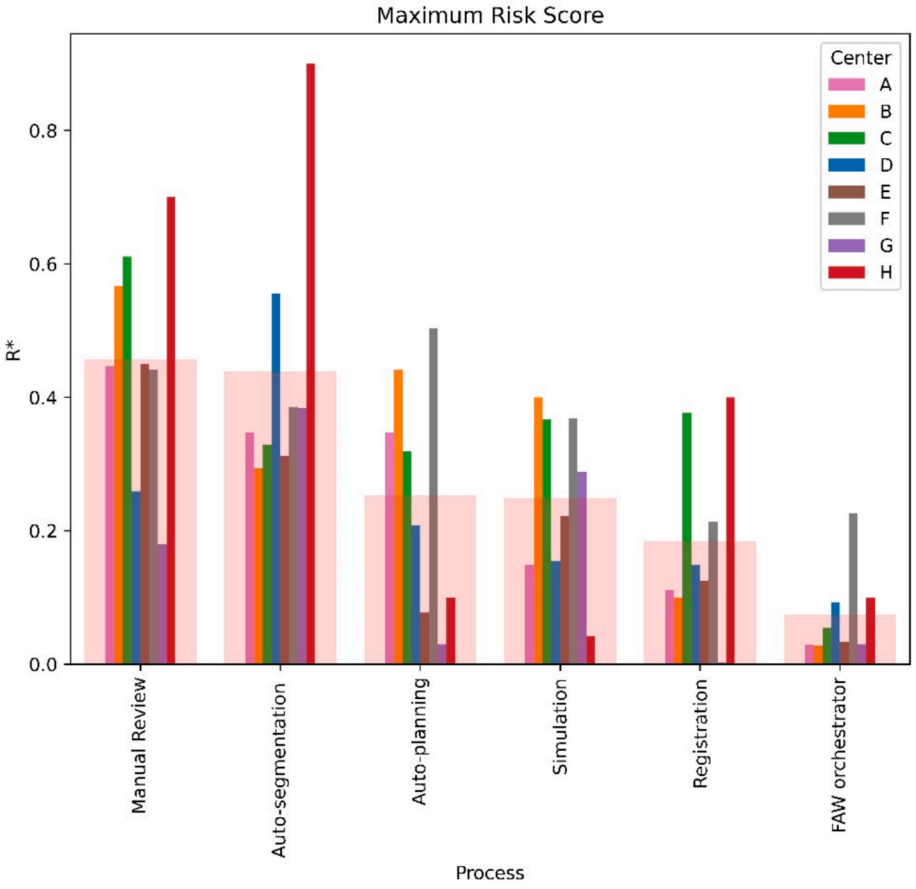

Fig. 2. Maximum normalized risk score R* per major FAW process step for every centre and averaged across all centres (light red bar plot). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

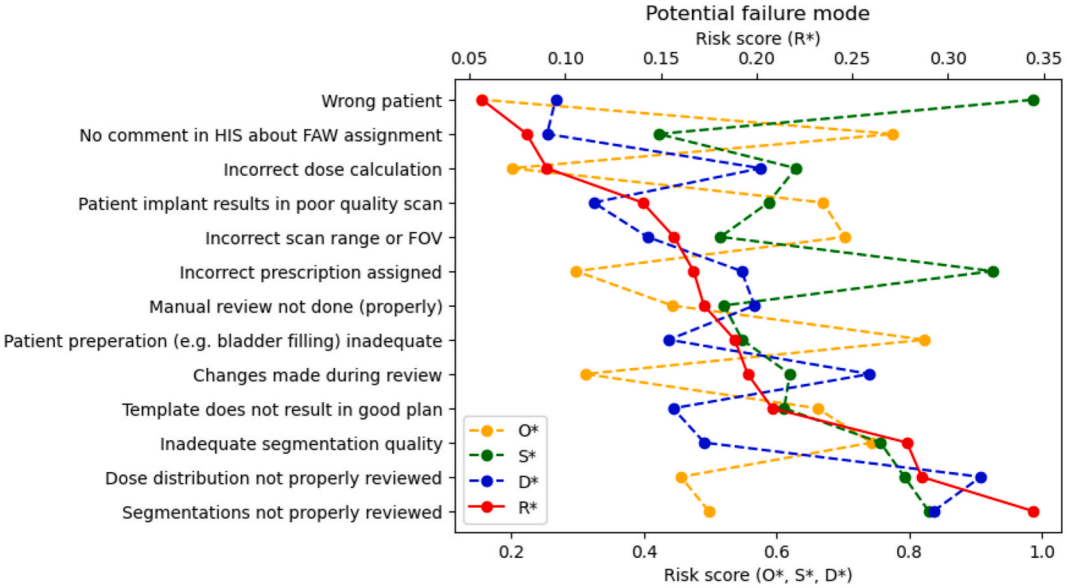

Fig. 3. Top five highest-scoring values for R*, O*, S* and D*and its corresponding failure mode, ordered by increasing R*. Due of overlap, the number of failure modes was limited to thirteen.

4. Discussion

This FMEA underlined the technical capability to implement FAWs in clinical routine. However, how people deal with automation was perceived as a high risk and tools and training are needed to support people reviewing the FAW output. Only when safe introduction is guaranteed, wide clinical adoption is feasible.

This multi-centre FMEA identified manual review as the highest risk process step, while the software architecture itself was scored as the lowest risk. This affirms the capability of software to automate

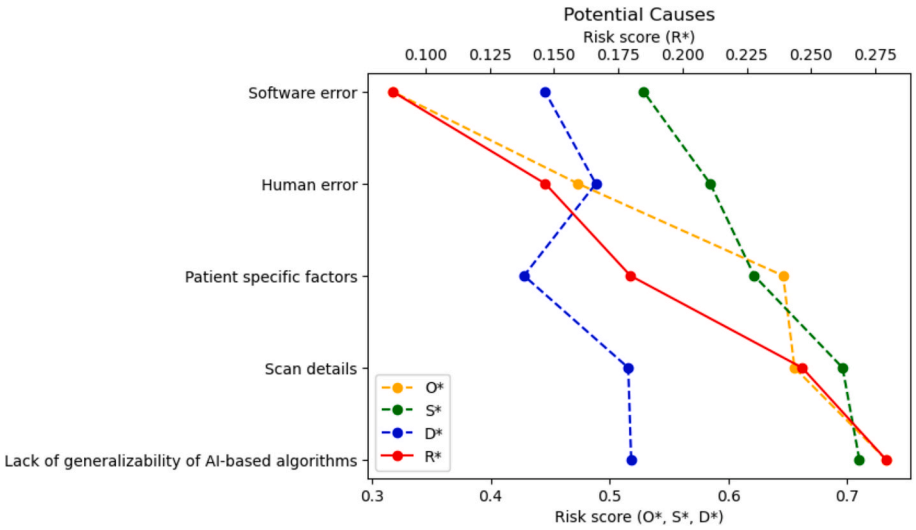

Fig. 4. R*, O*, S* and D* for every potential cause, ordered by increasing R*.

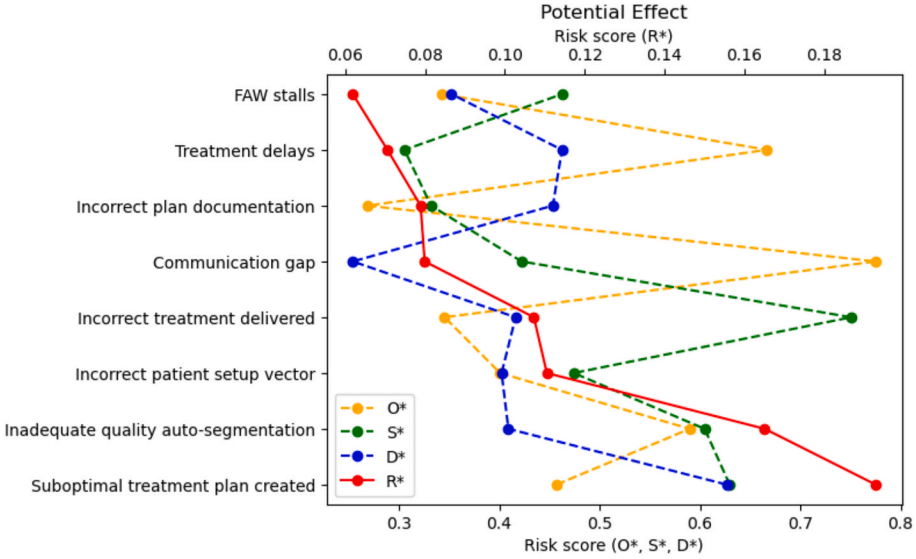

Fig. 5. R*, O*, S* and D* for every potential effect, ordered by increasing R*.

Table 1
Per major process step, the main message of the qualitative analysis is listed and the corresponding failure mode with the highest risk score.

Process	Qualitative summary	Quantitative summary
Patient assignment	Standardisation challenges and deviations are common (e.g. re-irradiation)	Incorrect protocol assigned
Auto-segmentation	Detecting and correcting subtle errors with impact becomes difficult with reduced human involvement (e.g. skill degradation)	Inadequate segmentation quality
Auto-planning	Detecting suboptimal plan will be challenging (e.g. skill degradation)	Template does not result in good plan
Manual review	Human oversight at the review stage is essential to catch mistakes. However, the risk of manual review not finding an error should not be underestimated.	Segmentation/dose distribution not properly reviewed

radiotherapy treatment preparation and aligns with previous findings that highlight the lack of trust, rather than the technical limitations, as the limiting factor for clinical adoption of FAWs [10,13]. Despite the potential high S*, technical software errors were expected to be rare because of a low O*. Care should be taken for software errors with high D*, as these errors might affect multiple patients or treatments.

Nevertheless, the auto-segmentation and auto-planning model's generalizability emerged as an important cause of concern. Especially when considered in combination with the inherent variability of individual patients and the practical impossibility to standardize every patient's treatment, FAW cannot guarantee the best output for every individual. Also within the initial, offline treatment preparation of an adaptive workflow, auto-segmentation and auto-planning were previously perceived as high-risk events [17]. This highlights the relevance of adequate manual verification, the need for new QA tools to assess FAW output, and supports the notion that careful consideration is required to define which workflows to automate and to what extent [24]. Applications of the FAW outside its intended use are therefore also recognized as

a key risk.

Large deviations are easy to detect, and minor deviations will have limited clinical impact, therefore "mid-level" errors need to be detected during manual review to prevent suboptimal treatment delivery, as demonstrated in the quantitative and qualitative analysis. However, this manual review step was identified as the highest risk in the entire workflow, strengthened by an automation-induced skill degradation and the risk of both automation and anchoring bias [27–29].

Auto-segmentation received a higher R* compared to the auto-planning, though variation in clinical expertise amongst centres might have influenced risk perception. Additionally, analysing a hypothetical workflow and its underlying assumptions is impactful and cannot be disregarded, which may be considered a limitation of this multi-centric study.

The failure modes with the highest S* were identified as assigning the wrong FAW protocol to a patient or vice versa. Despite the low D* and O*, avoiding these problems is key when using automated workflows. Proper training and understanding of the risks of the FAW by all clinical staff is essential to mitigate this risk.

To fully exploit the potential of FAW in radiation oncology, clinical adoption is essential [16]. This FMEA analysis reinforces the importance of 'quality by design' in AI model development and underscores the need of quality monitoring and education. Robustness of the FAW to protocol deviations might be poor, further highlighting the importance of protocol adherence. To overcome safety hurdles and to ensure standardization, checklists are often used [30,31]. However, because of the changing role of the medical staff, new FAW-specific and automated QA tools need to be developed. These QA tools should check the appropriateness of the FAW and assure that protocols are assigned correctly [32]. Moreover, QA tools should help to minimize the risks found in this analysis in the manual review. Monitoring tools should provide continuous feedback loops between clinical use and development to ensure safe and effective deployment, sensitive to clinically relevant changes [22,33,34].

The multi-centric study design and the explicit assumption to exclude QA or sanity checks can be seen as current limitations of this study. The multi-centric approach required data normalization for more reliable risk comparison between centres, but eliminated the possibility of an absolute risk assessment. Within the context of finding consensus, and by providing the raw data as Supplementary Material this study aimed to minimize the potential negative impact of the study design. The exclusion of QA might have an impact on the perceived risks, but as long as no guidelines exists on how to QA FAWs, adding QA to the hypothetical workflow would increase complexity and potentially also variation between centres. Already within the current study, large variation was present but the impact of used software, clinical expertise, and level of standardization on the risk scores was not within the scope of the current research question.

In the end, a new balance must be found between monitoring and trusting automation. The level of trust and required human interaction will depend on the complexity of the automation process. Therefore, every implemented FAW should be preceded by an implementation strategy, clear goals and a proper training of staff [20]. As automation will dramatically redefine the role of staff members, training will be needed on how to monitor and manage automation and how to deal with unforeseen circumstances [7,35].

The FMEA analysis highlighted wariness in the consistency of people dealing with automation rather than lack of trust in the FAW itself. Major concerns were the ability of people to correctly judge output in case of low generalizability of the model and increasing skill degradation. Consequently, tools and education are needed to help users interpret FAW output before large scale clinical application is feasible.

CRediT authorship contribution statement

Geert De Kerf: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Ana Barragán-Montero:** Conceptualization, Methodology, Writing – review & editing. **Charlotte L. Brouwer:** Conceptualization, Methodology, Writing – review & editing. **Pietro Pisciotta:** Conceptualization, Methodology, Writing – review & editing. **Marie-Claude Biston:** Conceptualization, Methodology, Writing – review & editing. **Marco Fusella:** Conceptualization, Methodology, Writing – review & editing. **Geoffroy Herbin:** Conceptualization, Methodology, Writing – review & editing. **Esther Kneepkens:** Conceptualization, Methodology, Writing – review & editing. **Livia Marrazzo:** Conceptualization, Methodology, Writing – review & editing. **Joshua Mason:** Conceptualization, Methodology, Writing – review & editing. **Camila Panduro Nielsen:** Conceptualization, Methodology, Writing – review & editing. **Koen Snijders:** Conceptualization, Methodology, Writing – review & editing. **Stephanie Tanadini-Lang:** Conceptualization, Methodology, Writing – review & editing. **Aude Vaandering:** Conceptualization, Methodology, Writing – review & editing. **Tomas M. Janssen:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

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

Acknowledgements

All participants to the FMEA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2025.100765.

References

[1] Franco P, Tesio V, Bertholet J, Gasnier A, Gonzalez del Portillo E, Spalek M, et al. Professional quality of life and burnout amongst radiation oncologists: the impact of alexithymia and empathy. Radiother Oncol 2020;147:162–8. https://doi.org/10.1016/j.radonc.2020.05.017.

[2] Hoffmans-Holtzer N, Smolenaers L, Olofsen-Van Acht M, Hoogeman M, Balvert M, Petit S. Robust optimization of a radiotherapy pretreatment preparation workflow. Phys Med Biol 2024;69:1–18. https://doi.org/10.1088/1361-6560/acf437.

[3] Thompson RF, Valdes G, Fuller CD, Carpenter CM, Morin O, Aneja S, et al. Artificial intelligence in radiation oncology: a specialty-wide disruptive transformation? Radiother Oncol 2018;129:421–6. https://doi.org/10.1016/j.radonc.2018.05.030. Artificial.

[4] Fu Y, Zhang H, Morris ED, Glide-Hurst CK, Pai S, Traverso A, et al. Artificial intelligence in radiation therapy. IEEE Trans Radiat Plasma Med Sci 2022;6:158–81. https://doi.org/10.1109/TRPMS.2021.3107454.

[5] Baroudi H, Brock KK, Cao W, Chen X, Chung C, Court LE, et al. Automated contouring and planning in radiation therapy: what is 'clinically acceptable'? Diagnostics 2023;13:1–21. https://doi.org/10.3390/diagnostics13040667.

[6] Kisling K, Johnson JL, Simonds H, Zhang L, Jhingran A, Beadle BM, et al. A risk assessment of automated treatment planning and recommendations for clinical deployment. Med Phys 2019;46:2567–74. https://doi.org/10.1002/mp.13552.

[7] Callens D, Malone C, Carver A, Fiandra C, Gooding MJ, Korreman SS, et al. Is full-automation in radiotherapy treatment planning ready for take off? Radiother Oncol 2024;201:110546. https://doi.org/10.1016/j.radonc.2024.110546.

[8] Simiele E, Han B, Skinner L, Pham D, Lewis J, Gensheimer M, et al. Mitigation of intensity modulated radiation therapy and stereotactic body radiation therapy treatment planning errors on the novel RefleXion X1 system using failure mode and effect analysis within six sigma framework. Adv Radiat Oncol 2023;8:1–10. https://doi.org/10.1016/j.adro.2023.101186.

[9] Nealon KA, Balter PA, Douglas RJ, Fullen DK, Nitsch PL, Olanrewaju AM, 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:e344–53. https://doi.org/10.1016/j.prro.2022.01.003.

[10] Nealon KA, Douglas RJ, Han EY, Kry SF, Reed VK, Simiele SJ, et al. Hazard testing to reduce risk in the development of automated planning tools. J Appl Clin Med Phys 2023;24:1–9. https://doi.org/10.1002/acm2.13995.

[11] Cavus H, Bulens P, Tournel K, Orlandini M, Jankelevitch A, Crijns W, et al. Safety and efficiency of a fully automatic workflow for auto-segmentation in radiotherapy using three commercially available deep learning-based applications. Phys Imaging Radiat Oncol 2024;31:100627. https://doi.org/10.1016/j.phro.2024.100627.

[12] Vandewinckele L, Benazzouz C, Delombaerde L, Pape L, Reynders T, Van der Vorst A, et al. Pro-active risk analysis of an in-house developed deep learning based autoplanning tool for breast volumetric modulated arc therapy. Phys Imaging Radiat Oncol 2024;32:100677. https://doi.org/10.1016/j.phro.2024.100677.
[13] Gooding MJ, Aluwini S, Guerrero Urbano T, McQuinlan Y, Om D, Staal FHE, et al. Fully automated radiotherapy treatment planning: a scan to plan challenge. Radiother Oncol 2024;200:110513. https://doi.org/10.1016/j.radonc.2024.110513.
[14] Vandewinckele L, Claessens M, Dinkla A, Brouwer C, Crijns W, Verellen D, et al. Overview of artificial intelligence-based applications in radiotherapy: recommendations for implementation and quality assurance. Radiother Oncol 2020;153:55–66. https://doi.org/10.1016/j.radonc.2020.09.008.
[15] Hussein M, Heijmen BJM, Verellen D, Nisbet A. Automation in intensity modulated radiotherapy treatment planning-a review of recent innovations. Br J Radiol 2018;91:20180270. https://doi.org/10.1259/bjr.20180270.
[16] McIntosh C, Conroy L, Tjong MC, Craig T, Bayley A, Catton C, et al. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. Nat Med 2021;27:999–1005. https://doi.org/10.1038/s41591-021-01359-w.
[17] Nishioka S, Okamoto H, Chiba T, Sakasai T, Okuma K, Kuwahara J, et al. Identifying risk characteristics using failure mode and effect analysis for risk management in online magnetic resonance-guided adaptive radiation therapy. Phys Imaging Radiat Oncol 2022;23:1–7. https://doi.org/10.1016/j.phro.2022.06.002.
[18] Klüter S, Schrenk O, Renkamp CK, Gliessmann S, Kress M, Debus J, et al. A practical implementation of risk management for the clinical introduction of online adaptive magnetic resonance-guided radiotherapy. Phys Imaging Radiat Oncol 2021;17:53–7. https://doi.org/10.1016/j.phro.2020.12.005.
[19] Huynh E, Hosny A, Guthier C, Bitterman DS, Petit SF, Haas-Kogan DA, et al. Artificial intelligence in radiation oncology. Nat Rev Clin Oncol 2020;17:771–81. https://doi.org/10.1038/s41571-020-0417-8.
[20] Swart RR, Jacobs MJG, Roumen C, Houben RMA, Koetsveld F, Boersma LJ. Factors predicting timely implementation of radiotherapy innovations: the first model. Br J Radiol 2021;94:1–9. https://doi.org/10.1259/bjr.20200613.
[21] Petragallo R, Bardach N, Ramirez E, Lamb JM. Barriers and facilitators to clinical implementation of radiotherapy treatment planning automation: a survey study of medical dosimetrists. J Appl Clin Med Phys 2022;23:1–10. https://doi.org/10.1002/acm2.13568.
[22] Liu X, Glocker B, McCradden MM, Ghassemi M, Denniston AK, Oakden-Rayner L. The medical algorithmic audit. Lancet Digit Heal 2022;4:e384–97. https://doi.org/10.1016/S2589-7500(22)00003-6.
[23] Cardoso MJ, Moosbauer J, Cook TS, Erdal BS, Genereaux B, Gupta V, et al. RAISE – Radiology AI Safety, an End-to-end lifecycle approach 2023. https://doi.org/10.48550/arXiv.2311.14570.
[24] Zayas-Cabán T, Haque SN, Kemper N. Identifying opportunities for workflow automation in health care: lessons learned from other industries. Appl Clin Inform 2021;12:686–97. https://doi.org/10.1055/s-0041-1731744.
[25] Liu HC, Zhang LJ, Ping YJ, Wang L. Failure mode and effects analysis for proactive healthcare risk evaluation: a systematic literature review. J Eval Clin Pract 2020;26:1320–37. https://doi.org/10.1111/jep.13317.
[26] Huq MS, Fraass BA, Dunscombe PB, Gibbons JP, Ibbott GS, Mundt AJ, et al. The report of Task Group 100 of the AAPM: application of risk analysis methods to radiation therapy quality management. Med Phys 2016;43:4209–62. https://doi.org/10.1118/1.4947547.
[27] Macnamara BN, Berber I, Çavuşoğlu MC, Krupinski EA, Nallapareddy N, Nelson NE, et al. Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? Cogn Res Princ Implic 2024;9:1–9. https://doi.org/10.1186/s41235-024-00572-8.
[28] Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Informatics Assoc 2012;19:121–7. https://doi.org/10.1136/amiajnl-2011-000089.
[29] Mackay K, Taylor A, Banfill K, Bernstein D, Daniel J, Diez P, et al. Clinical Oncology Guidance on auto-contouring in radiotherapy. London: 2024. https://www.rcr.ac.uk/media/rqjlnlny/rcr-auto-contouring-in-radiotherapy-2024.pdf.
[30] Rassiah P, Su FCF, Huang YJ, Spitznagel D, Sarkar V, Szegedi MW, et al. Using failure mode and effects analysis (FMEA) to generate an initial plan check checklist for improved safety in radiation treatment. J Appl Clin Med Phys 2020;21:83–91. https://doi.org/10.1002/acm2.12918.
[31] Lastrucci A, Esposito M, Serventi E, Marrazzo L, Francolini G, Simontacchi G, et al. Enhancing patient safety in radiotherapy: Implementation of a customized electronic checklist for radiation therapists. Tech Innov Patient Support Radiat Oncol 2024;31:100255. https://doi.org/10.1016/j.tipsro.2024.100255.
[32] O'Daniel J, Masgrau VH, Clark C, Esposito M, Lehmann J, McNiven A, et al. Which failures do patient-specific quality assurance systems need to catch? Med Phys 2024;52:88–98. https://doi.org/10.1002/mp.17468.
[33] De Kerf G, Claessens M, Raouassi F, Mercier C, Stas D, Ost P, et al. A geometry and dose-volume based performance monitoring of artificial intelligence models in radiotherapy treatment planning for prostate cancer. Phys Imaging Radiat Oncol 2023;28:100494. https://doi.org/10.1016/j.phro.2023.100494.
[34] Barragan-Montero A, Bibal A, Dastarac MH, Draguet C, Valdes G, Nguyen D, et al. Towards a safe and efficient clinical implementation of machine learning in radiation oncology by exploring model interpretability, explainability and data-model dependency. Phys Med Biol 2022;67:1–46. https://doi.org/10.1088/1361-6560/ac678a.
[35] Hallows R, Glazier L, Katz MS, Aznar M, Williams M. Safe and ethical artificial intelligence in radiotherapy – Lessons learned from the aviation industry. Clin Oncol 2022;34:99–101. https://doi.org/10.1016/j.clon.2021.11.019.