ACTA ONCOLOGICA
2026, VOL. 65, 652–654
https://doi.org/10.2340/1651-226X.2026.46249

NACP 2026

MJS PUBLISHING

EDITORIAL

Medical physics in the Nordic countries: collaboration and future challenges – report from the Nordic Association for Clinical Physics 2026 Symposium

Emelie Adolfsson[a,b] and Jan Seppälä[c]

[a]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden; [b]Clinical Department of Medical Radiation Physics, Region Östergötland, Linköping, Sweden; [c]Department of Radiotherapy, Kuopio University Hospital, Kuopio, Finland

ARTICLE HISTORY
Received 3 June 2026
Accepted 20 June 2026
Published 30 July 2026

The Nordic Association for Clinical Physics (NACP) held its triennial symposium in Gothenburg, Sweden, from 24 to 26 March 2026. Under the theme 'Collaboration and future challenges', the symposium brought together medical physicists from radio-therapy, radiology, nuclear medicine, and magnetic resonance, providing a comprehensive overview of emerging technologies and their translation into clinical practice in the Nordic region. The symposium comprised plenary sessions, parallel scientific sessions, workshops, and poster presentations, offering a platform for both emerging research and clinical implementation. A recurring message throughout the symposium was the role of the medical physicist not only as a physics expert but also as a key contributor to safe and effective patient care. The meeting captured both scientific advances and practical challenges facing medical physicists today.

The symposium opened with an engaging retrospective on the history of NACP, founded in 1962, presented by the current president, Toni Ihalainen (FI). He highlighted the association's early international impact, including contributions to developments that later formed international standards such as the International Atomic Energy Agency (IAEA) dosimetry protocol. Following a 12-year period of inactivity, the association was revitalized with the approval of a new constitution by the boards of the Nordic national member organizations in 2004. For physicists whose professional experience is largely rooted in the 21st century, this historical perspective provided a valuable and, in many ways, eye-opening insight into the legacy and evolution of the field.

A recurring theme of the two most recent NACP symposia has been the introduction of artificial intelligence (AI) in medical physics, and the discussions at the 2026 symposium reflected a clear maturation of the field. Whereas earlier meetings emphasized methodological development, the 2026 symposium shifted the focus toward clinical implementation and practical deployment. This transition was exemplified in plenary lectures by Mika Kortesniemi (FI), 'The role of the Medical Physicist with respect to AI', and Mark Gooding, 'RoboRadOnc: The future of radiation therapy?'. Together, these presentations underscored both the evolving role of the medical physicist and the broader trajectory of the field, pointing toward increasingly data-driven, clinically grounded, and automated approaches to patient care.

Scientific themes emerging from NACP 2026

Scientific contributions from proffered papers and posters reflected a broad range of topics, including predictive modeling of treatment-related toxicity, ultra-high dose rate radiotherapy, and target motion management.

Reirradiation is becoming an increasingly relevant component of radiotherapy, driven by several converging developments. Improved cancer survival has led to a growing population of patients presenting with local recurrences or new primary tumors in previously irradiated regions. At the same time, advances in treatment delivery, imaging, and dose planning have enabled more precise approaches, allowing reirradiation to be considered in situations where it was previously deemed too high risk. In this context, the plenary contribution by Ane Appelt was particularly timely. As a key figure in shaping international standards and research agendas in reirradiation, her presentation provided a comprehensive overview of current clinical practice alongside ongoing research efforts, highlighting both the opportunities and the remaining challenges in this evolving area [1].

A notable development is the growing emphasis on standardization and multi-institutional collaboration. While data-driven approaches have been discussed in earlier meetings [2], the 2026 symposium placed greater focus on structured clinical data, national registries, and coordinated efforts to support consistent practice. The KaSARi (Key advances in Standardizing, Automating, and predicting Risks in radiotherapy) initiative in Finland represents a particularly ambitious example of this development. Unlike a conventional

CONTACT Emelie Adolfsson emelie.adolfsson@regionostergotland.se Department of Medical Radiation Physics, Linköping University Hospital, entrance 34, S-581 85 Linköping, Sweden

retrospective registry, KaSARi is designed as a prospective national infrastructure integrated into routine clinical workflows, linking DICOM-RT data with toxicity and patient-reported outcomes [3]. The framework aims to provide a sustainable foundation for nationwide outcome evaluation, harmonization of radiotherapy practices, and future development of predictive and AI-based methods. In Denmark, a national collaboration has focused on patient safety through systematic analysis of adverse events, resulting in an updated framework for classification and risk assessment of unintended incidents in radiotherapy [4]. Beyond improving consistency in reporting, the initiative provides a mechanism for shared learning across institutions and illustrates how quality improvement increasingly relies on structured national collaboration rather than local experience alone. These efforts illustrate how future progress in medical physics will depend on robust infrastructures for collecting, standardizing, and actively using clinical data in practice.

As radiotherapy becomes increasingly advanced and data driven, greater attention must be paid not only to technical capabilities but also to their limitations, uncertainties, and clinical consequences. Several contributions highlighted the increasing complexity of modern radiotherapy and the need to evaluate both performance and robustness. Advances in MRI-based segmentation demonstrated that future clinical adoption of AI depends not only on geometric accuracy but also on the ability to identify when predictions are uncertain [5]. The study showed that improved uncertainty assessment and calibration can increase confidence in automated contouring by highlighting regions where manual review may be warranted, thereby supporting safer integration of AI into clinical radiotherapy workflows. A similar focus on uncertainty extends to other parts of the treatment chain. Differences between dose calculation algorithms become increasingly relevant as treatment plans grow more complex, underscoring the importance of understanding algorithm-specific limitations when evaluating highly modulated techniques [6]. In parallel, a comparison of volumetric modulated arc therapy and three-dimensional conformal radiotherapy for whole-breast treatment including simultaneous boost demonstrated improved dose conformity with VMAT but also illustrated trade-offs in organ-at-risk doses, emphasizing that more advanced techniques do not automatically translate into overall clinical benefit and must be evaluated in terms of clinical balance [7].

The interplay between complexity and reliability was also evident in quality assurance and dosimetry studies. An evaluation of starshot methods for isocenter size determination showed that EPID-based approaches can achieve comparable accuracy to traditional film-based approaches. Importantly, the work also demonstrated that the choice of beam collimation influenced the estimated isocenter size, highlighting how quality assurance results may depend not only on machine performance but also on the methodology used for assessment [8]. In ophthalmic brachytherapy using ^{125}I seeds, accurate dose verification remains challenging because of the steep dose gradients and short source-to-detector distances involved [9]. These findings serve as a reminder that improvements in treatment precision must be matched by equally rigorous approaches to dosimetry and quality assurance. In highly localized treatments such as ophthalmic brachytherapy, even small measurement uncertainties may have a substantial impact on dose verification. Efforts toward more individualized treatment approaches were also reflected in radionuclide therapy, where pre-treatment prostate-specific membrane antigen positron emission tomography imaging was explored as a means of estimating absorbed doses during subsequent ^{177}Lu PSMA therapy [10]. While substantial variability remained, particularly for tumor lesions, the results illustrate ongoing efforts to move toward patient-specific treatment planning in molecular radiotherapy.

Proton therapy has featured prominently at previous NACP meetings [11–13], yet in 2026 only two contributions were presented: clinical implementation of a dual-energy CT workflow [14] and the Norwegian Proton and Radiotherapy Registry [15]. With proton therapy centers now operational in Sweden, Denmark, and Norway, one might have expected a wider range of collaborative projects and data-sharing initiatives to be represented. The relatively limited representation of proton therapy may indicate that many ongoing collaborative activities are now being communicated through dedicated proton therapy networks, reflecting the maturation of the field within the Nordic region.

Summary

In summary, the radiotherapy contributions at NACP 2026 reflect a field in transition. The key challenges are shifting from the development of technologically advanced solutions to their implementation, evaluation, and refinement in clinical practice. This includes systematic learning from previous treatments, structured follow-up of patient outcomes, and the growing use of clinical data to guide decision-making. The increasing integration of data-driven approaches and AI-based models further supports the move toward more individualized and adaptive treatment strategies. In this context, the role of medical physicists is evolving toward a broader clinical and integrative role. This development places new demands on the profession, not only in clinical practice but also in education and training, which must adapt to prepare future physicists for these expanded responsibilities. At the same time, it remains essential that this broader role is grounded in the core principles of medical physics and radiation science.

Acknowledgements

NACP 2026 was financially supported by the Acta Oncologica Foundation.

Conflicts of interest

The authors report there are no competing interests to declare.

Data availability statement

N/A.

References

[1] Paradis KC, Appelt AL, Bentzen SM, Chng N, Covington EL, Dawson LA, et al. Reirradiation Collaborative Group (ReCOG) consensus on standards for dose evaluation and reporting in patients with multiple courses of radiation therapy: an AAPM/ACRO/ASTRO/CARO/COMP/CADRA/CPQR/ESTRO/NRG-endorsed consensus statement. Lancet Oncol. 2026;27(2):e101–10. https://doi.org/10.1016/S1470-2045(25)00562-5

[2] Long Krogh S, Brink C, Laugaard Lorenzen E, Samsøe E, Richter Vogelius I, Zukauskaite R, et al. A national repository of complete radiotherapy plans: design, results, and experiences. Acta Oncol. 2023;62(10):1161–8. https://doi.org/10.1080/0284186X.2023.2270143

[3] Seppälä J, Tengström M, Korkalainen H, Virén T, Honkanen JTJ, Nikkinen J, et al. KaSARi: a national framework for standardized, automated, and predictive radiotherapy in Finland. Acta Oncol. 2026;65:430–6. https://doi.org/10.2340/1651-226X.2026.45682

[4] Jakobsen KL, Jakobsen AR, Gottlieb KL, Berg M, Spejlborg H, Biancardo SBN, et al. An updated system for categorising and reporting unintended incidents in radiotherapy at a national level. Acta Oncol. 2026;65:416–21. https://doi.org/10.2340/1651-226X.2026.45699

[5] Kallehauge JF, Ren J, Lassen-Ramshad Y. Improving confidence in MRI-based auto-segmentation via uncertainty assessment. Acta Oncol. 2026;65:406–11. https://doi.org/10.2340/1651-226X.2026.45685

[6] Terzidis E, Nordström F, Gustafsson M, Karlsson A, Götstedt J, Bäck A. Dose calculation errors for volumetric modulated arc therapy plans of various complexity. Acta Oncol. 2026;65: 614–623. https://doi.org/10.2340/1651-226X.2026.45698

[7] Johansson G, Kojoj R, Surdashi S, Olin M, Fredén E, Sen P, et al. Evaluation of 3D-CRT vs. VMAT for radiotherapy of whole breast with simultaneous integrated boost. Acta Oncol. 2026;65: 543–552. https://doi.org/10.2340/1651-226X.2026.45677

[8] Honkanen A-P, Kulmala A. Starshot field configuration in the linac isocentre size determination: a comparison of film and electronic portal imaging device measurements. Acta Oncol. 2026;65: 557-560. https://doi.org/10.2340/1651-226X.2026.45692

[9] Rintala A, Reijonen V, Valve A, Tenhunen M. Validation of absolute dose calculation of a single 125I-seed for ophthalmic brachytherapy within distance range 1–15 mm using a diode detector. Acta Oncol. 2026;65:511–7. https://doi.org/10.2340/1651-226X.2026.45688

[10] Kokkonen S, Reijonen V, Hippeläinen E, Ahtiainen V, Schildt J, Sipilä O, et al. Preliminary feasibility of pre-treatment F-18-PSMA-1007 PET/CT in dose prediction for Lu-177-PSMA-I&T therapy. Acta Oncol. 2026;65: 609–613. https://doi.org/10.2340/1651-226X.2026.45700

[11] Rønde HS, Kallehauge JF, Kronborg CJS, Nyvang L, Rekstad BL, Hanekamp BA, et al. Intensity modulated proton therapy planning study for organ at risk sparing in rectal cancer re-irradiation. Acta Oncol. 2021;60(11):1436–9. https://doi.org/10.1080/0284186X.2021.1953139

[12] Brincker M, Jensen I, Ann Rechner L, Anne Schut D, Stagaard Johansen T, Nielsen M, et al. Multi-center comparison between proton and photon plans for mediastinal lymphomas. Acta Oncol. 2023;62(10):1251–5. https://doi.org/10.1080/0284186X.2023.2251089

[13] Rønde HS, Kronborg C, Høyer M, Hansen J, Eppler Bak M, Nielsen Agergaard S, et al. Dose comparison of robustly optimized intensity modulated proton therapy (IMPT) vs IMRT and VMAT photon plans for testicular seminoma. Acta Oncol. 2023;62(10):1222–9. https://doi.org/10.1080/0284186X.2023.2254925

[14] Pettersson E, Thilander Klang A, Gillies C, Clarke M, Bäck A. Assessing accuracy and artefacts in proton stopping power ratio images across four computed tomography imaging workflows using a head-sized electron density phantom. Radiat Prot Dosimetry. 2026;202(3–4):381–92. https://doi.org/10.1093/rpd/ncaf159

[15] Pettersen HES. Norwegian Proton and Radiotherapy Registry (NORPREG). Helse Bergen HF; 2026 [cited 2026 June 01]. Available from: https://www.helse-bergen.no/avdelinger/kreftklinikken/flerregional-behandlingstjeneste-for-protonterapi/kvalitetsverktoy/

ACTA ONCOLOGICA
2026, VOL. 65, 416–421
https://doi.org/10.2340/1651-226X.2026.45699

NACP 2026

MJS PUBLISHING

ORIGINAL ARTICLE

An updated system for categorising and reporting unintended incidents in radiotherapy at a national level

Kirsten L. Jakobsen[a], Annette R. Jakobsen[b], Karina L. Gottlieb[c], Martin Berg[d], Harald Spejlborg[e], Susan B.N. Biancardo[f], Bob Smulders[g] and Heidi S. Rønde[h]

[a]Department of Oncology and Palliative Care, University Hospital Region Zeeland, Næstved, Denmark; [b]Department of Medical Physics, Department of Oncology, Aalborg University Hospital, Aalborg, Denmark; [c]Laboratory of Radiation Physics, Department of Oncology, Odense University Hospital, Odense, Denmark; [d]Department of Medical Physics, Department of Oncology, Vejle Hospital, Lillebaelt Hospital – University Hospital of Southern Denmark, Vejle, Denmark; [e]Department of Medical Physics, Aarhus University Hospital, Aarhus, Denmark; [f]Department of Oncology, Copenhagen University Hospital – Herlev and Gentofte, Denmark; [g]Department of Oncology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; [h]Danish Centre for Particle Therapy, Aarhus University Hospital, Aarhus, Denmark

ABSTRACT

Background and purpose: Radiotherapy is a complex, multistage process that involves collaboration across professional disciplines (radiographers working in pre-treatment preparation, physicians, physicists, radiation therapists during planning and treatment. While severe errors are uncommon, they can have profound consequences for patients. The Danish Society for Medical Physics supports a Special Interest Group dedicated to improving patient safety by analysing unintended events (UEs) and near misses in radiotherapy. Here we report on the work conducted with an updated system for categorising UEs in radiotherapy in Denmark.

Material and methods: An updated reporting system based on the English Radiotherapy Pathway Coding (RPC) system has been taken into use in 2025. This system, including the use of a Danish system that ranks the UEs from the patients perspective, called Patient Centred Coding, is outlined and UEs reported in 2025 are summarised.

Results: Radiotherapy is generally a safe treatment modality, a conclusion further supported by our results from 2025. In 2025 only 339 UEs were reported, 315 of these with low risk and 24 with medium risks. The lesson learned is that the main focus in order to prevent the most harmful UEs should be on the delineation and the planning process.

Interpretation: Based on our experience from this work, we recommend the development of a national system for categorising and ranking UEs. We further recommend establishing a national working group that meets regularly and focuses on UEs at a national level.

ARTICLE HISTORY
Received 10 March 2026
Accepted 24 April 2026
Published 11 May 2026

KEYWORDS
Radiotherapy; patient safety; unintended events; risk management

Introduction

Radiotherapy is one of the most widely applied treatment modalities for cancer both as a primary intervention and as adjuvant therapy. In Denmark approximately 47,000 individuals are diagnosed with cancer every year and an estimated 50% receive at least one course of radiotherapy [1]. Despite its complexity and the involvement of multiple professional disciplines (radiographers (pre-treatment preparation), physicians, physicists and radiation therapists (RTTs), radiotherapy is generally regarded as a safe treatment modality, with a low number of reported unintended events (UEs) [2]. While most UEs in radiotherapy have minor effects and do no harm to the patients involved, there still may remain a risk for UEs leading to serious patient consequences. Especially those UEs related to, for example, malfunctions of equipment and systematic human errors can potentially affect the clinical outcome of numerous patients.

Given these considerations, systematic efforts to ensure patient safety are essential. In Denmark, this work is supported by the 'Danish Patient Safety Database' (DPSD) [3], which provides an established infrastructure for reporting UEs across the entire healthcare system in general as confirmed in the Danish health act [4]. However, DPSD is a general healthcare reporting system, which does not let users categorise UEs into radiotherapy-specific sub-categories. This makes systematic monitoring, reporting and comparisons within radiotherapy difficult. Thus, analysis of patient safety within Danish radiotherapy calls for an additional system for categorisation and description of UEs. National systematic reporting enables continuous monitoring of safety issues across Danish

CONTACT Kirsten L. Jakobsen kijak@regionsjaelland.dk Rådmandsengen 5, 4700 Næstved, Denmark
Supplemental data for this article can be accessed online at https://doi.org/10.2340/1651-226X.2026.45699

radiotherapy centres and creates the empirical basis necessary for analysing error patterns. This approach allows for the identification of system vulnerabilities and facilitates the implementation of targeted preventive measures. Strengthening the reporting culture and ensuring consistent documentation of UEs are therefore crucial steps towards reducing the frequency of such events and maintaining a high level of patient safety in radiotherapy.

In 2012, the Danish Cancer Society established a working group to evaluate the safety of radiotherapy nationally. A literature review and international comparison highlighted the Radiotherapy Pathway Coding system (RPC) developed in the United Kingdom [5], which was subsequently adapted for use in the Danish context. Patient representatives participating in the working group identified a need to complement the RPC system with a Patient Centered Code (PCC) framework.

As a direct outcome of the 2012 national meeting, a Special Interest Group (SIG) on Patient Safety in radiotherapy was established under the auspices of the Danish Society for Medical Physics (DSMF) in 2017. Although case sharing continued, the group's activity over the subsequent years was intermittent, with limited engagement and irregular, informally structured meetings. In 2022, the remaining active members undertook a structured revitalisation of the initiative, formalising its governance, establishing terms of reference, and reconstituting the SIG into its current form. Since the SIG was established under DSMF (whose members comprise solely physicists) and today still reports to the DSMF board, the group has historically not been multidisciplinary. Even so, the SIG has now been expanded to include both RTTs, radiographers and medical physicists; a radiation oncologist representative has not yet been recruited to the national group although physicians are active in all local groups. As documented here there is a long tradition and legal requirement in Denmark for reporting UEs. In the later years the importance of this is recognised globally and more and more countries are working in the same direction [6–9].

The SIG has developed an updated RPC system tailored to current radiotherapy-practices in Denmark. The result of this work is presented in this article along with the PCC scoring exemplified by the UEs from 2025.

Material and methods

DPSD is a system that surveys the safety of the total healthcare system in Denmark. Employees are obliged to report both actual and potential (near misses) serious and lethal UEs, including UEs with a learning potential. This can be done anonymously. From DPSD the case is transferred to the specific department of origin. Each radiotherapy centre has a multidisciplinary panel that reviews the UEs and ensures learning from relevant cases. Importantly, the DPSD is a non-sanctioning environment.

A UE report in DPSD contains a free text description and a risk evaluation. Thus, all UEs and near misses are registered with an actual consequence and a potential consequence. These are then merged to the combined risk of the event (Table 1).

Table 1. Risk scoring when reporting UEs in DPSD.

Actual consequence		Potential consequence		Combined risk
None/unknown	+	None	=	Low
None/unknown	+	Minor/Moderate	=	Low
None/unknown	+	Serious	=	Medium
None/unknown	+	Lethal	=	High
Minor/moderate	+	Minor/moderate	=	Low
Minor/moderate	+	Serious	=	Medium
Minor/moderate	+	Lethal	=	High
Serious	+	Serious	=	High
Serious	+	Lethal	=	High
Lethal	+	N/A	=	High

The first and second columns refer to the actual and potential consequence of the UE under consideration. The combination of these two leads into a combined risk shown in the third column. UE: unintended events; DPSD: Danish Patient Safety Database.

While reviewing the UE categories provided in the RPC system, it became evident that the original set of 21 UE origins was no longer up-to-date. A closer analysis of the current radiotherapy workflow resulted in an updated RPC system with 14 categories (Table 2). These range from room design to handling of side effects. All 14 categories have been divided into sub-categories giving a total of 150 sub-categories (See Supplementary Material). The subdivision enables a more precise analysis of where in the processes UEs occur and focuses the work towards preventing UEs even more. Naturally, such subgroup analyses are only feasible when sufficient data are available to ensure statistical significance.

The PCC scores the potential risk for the patient, with low scores corresponding to high perceived risk. The most severe effect for a patient would be an impact of the UE on the treated volume. This would affect the tumour control probability, risk of side effects, or both. A UE affecting the follow-up would be of less harm as seen from the patient perspective, leading to a higher PCC score. The framework of the PCC is outlined in Table 3.

Table 2. The Danish RPC-coding anno 2025 consists of 14 items.

Process code	Activity description
0	Room design
1	External visitation and referral
2	Internal visitation and referral
3	Doctors' appointment
4	Booking process (pretreatment and treatment)
5	Fixation before and at scanner
6	Scanning
7	Image handling
8	Delineation
9	Dose planning
10	Independent checks
11	Treatment
12	Medication/concomitant treatment
13	Handling of side effects

RPC: Radiotherapy Pathway Coding.

Table 3. The Danish patient centered code (PCC) system consists of 10 items.

Item	Impact area
1	Treatment volume
2	Dose/fractionation
3	Waiting time/treatment delay
4	Concomitant treatment
5	Technical equipment
6	Independent checks
7	Communication/documentation
8	Patient experience communication/documentation
9	Follow-up
10	Other

The lower the number, the greater the effect on the patient as seen from the patient's perspective.

Results

Anonymised UE data including PCC and RPC, patient actual and potential harm, and combined overall risk from each of the eight Danish radiotherapy centres reported in 2025, where more than 240,000 external radiotherapy fractions were given, were collected. In total 339 UEs were reported: 315 incidents with low and 24 incidents with medium combined patient risk. This is comparable to the numbers in 2024: 315 UEs, 299 incidents with low risk, 15 with medium risk and 1 with high risk. As we have changed the RPC system the categorisation cannot be compared directly between the 2 years.

Figure 1 shows the distribution of UEs across the various steps in the radiotherapy workflow. Events most frequently (27%) originate in the treatment process, which is likely due to the multiple number of fractions compared to the single instance of planning and simulation stages. However, a considerable number of UEs are also reported during scanning (13%), delineation (9%) and treatment planning (12%).

While Figure 1 demonstrates in which workflow steps there may be room for improvements, it yields no information on the severity of the UEs as seen from a patient perspective. This information is provided by the PCC data. Figure 2 shows the ranking of the UEs in the PCC system. It is seen that the largest number of UEs (n = 83, 24%) is found in category 1 (treatment volume), which are also the UEs carrying the highest risk seen from the patients' point of view. In order to find the RPC categories where UEs with the most harmful potential originate, all UEs with PCC category 1 were categorised by the RPC system. Figure 3 demonstrates that most UEs affecting treatment volume originate in delineation (n = 27, 33%) and treatment (n = 25, 30%).

Discussion and conclusion

Learning from reported UEs is essential within radiotherapy, where errors can lead to serious harm to patients. A prerequisite to learning is a system supporting the reporting itself, a meaningful categorisation of the UEs, and last but not least presentation of the collected data to all staff members contributing to the radiotherapy course. In Denmark the reporting system was already present, but a standardised, contemporary national radiotherapy-specific system for categorising UEs was missing. This has now been developed and implemented.

During the analysis of the UEs in 2025 the SIG had access to UEs on an institutional level. In order to maintain the anonymity of the centres only the collective results are presented in this paper.

The UEs and their distribution patterns were found to be very similar across the eight Danish departments, even though both their software (Oncology Information Systems, Treatment

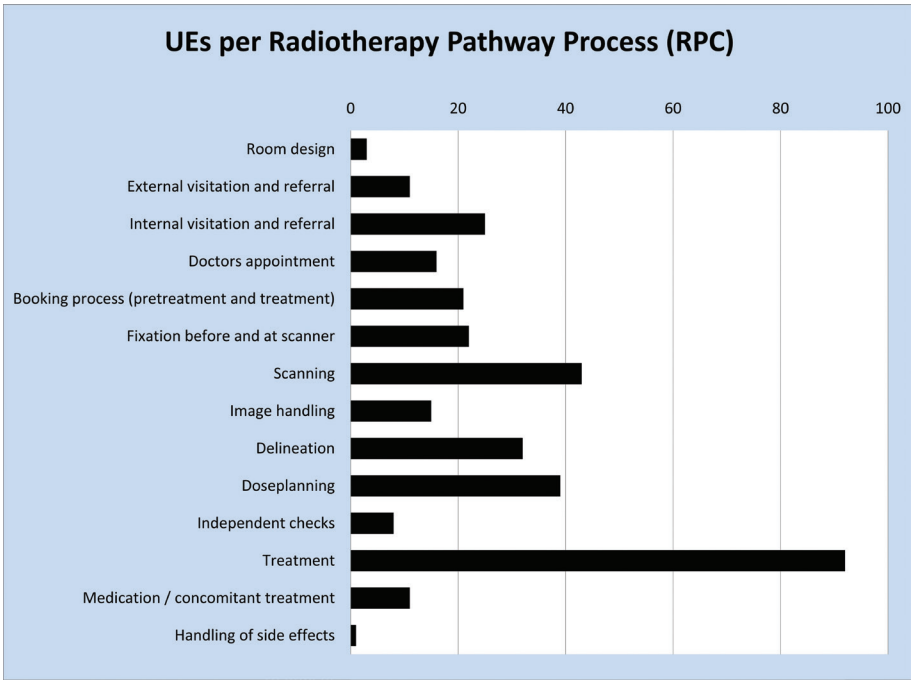


Figure 1. Unintended events categorised in RPC. RPC: Radiotherapy Pathway Coding.

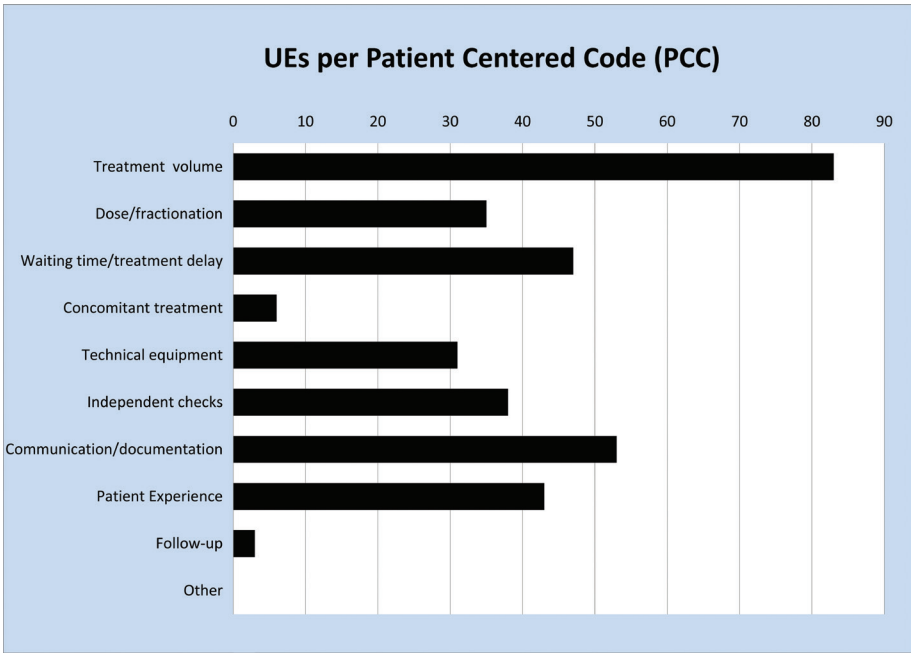


Figure 2. Unintended events categorised in PCC. PCC: patient centered code.

Planning Systems) and hardware (computer tomography [CT], magnetic resonance imaging [MRI], linear accelerators) differ. This was very encouraging since it supports the idea that clinics can learn from each other in spite of differences in their technological set-up. The PCC and updated RPC categorisations generate more detailed knowledge on where in the radiotherapy process UEs occur and how severe they are. This makes it possible to focus efforts where it yields the highest value, for example, on multiple similar UEs or UEs with high impact on patient safety.

No UEs with a high risk score were reported, indicating that no events with serious or lethal consequences occurred in 2025. This suggests that radiotherapy in Denmark remains a safe treatment modality. The error rate of 1.4 per 1000 treatments is comparable to the rate of 1.3 per 1000 found by Smith et al. [10].

Based on the reported distribution of UEs (Figures 1 and 2) it seems relevant to focus efforts around the pre-treatment workflow steps; delineation, prescription and dose planning. Errors in these processes may systematically influence the entire

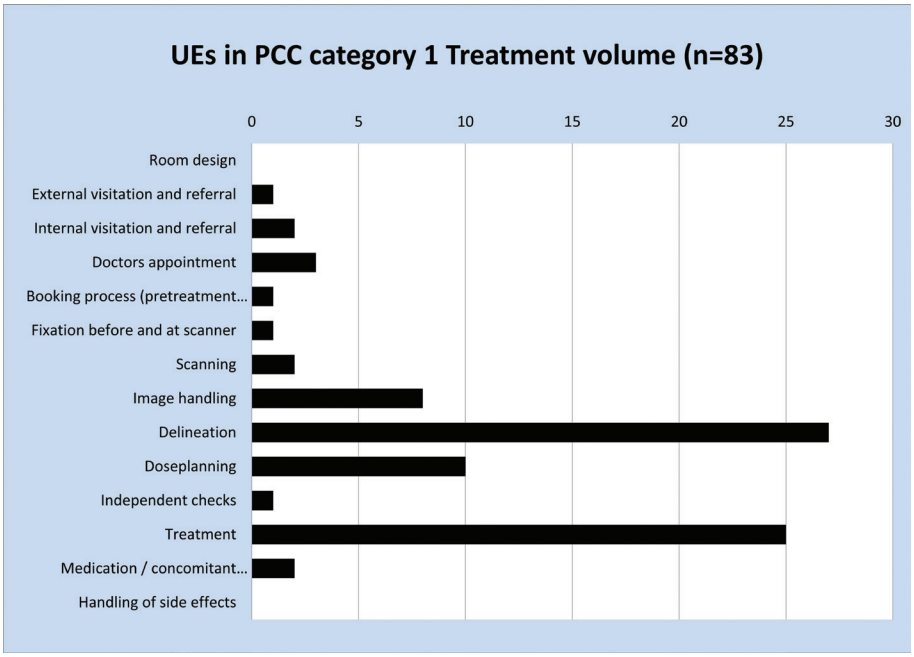


Figure 3. Origin in the RPC system for unintended events in PCC category 1 (Treatment volume). PCC: patient centered code; RPC: Radiotherapy Pathway Coding.

treatment, as opposed to an error occurring only once in the course of multiple treatment fractions. By having a national system to report UEs, improvements are based on statistics, and not 'feelings', and a common language has evolved. Experience has shown that it is important to maintain regular meetings with participants from across the country; otherwise these common agreements tend to evolve in different directions.

One of the limitations of any reporting system is the risk of under-reporting of UEs. In Denmark this work is supported by the law [11]. This means local management supports allocating resources to the task of collecting and disseminating the data. Furthermore, a non- sanctioning system with optional anonymous reporting promotes higher reporting rates and leads to stronger organisational learning. A system built as described in this work is known to increase the numbers of reported UEs [12].

In general, there is an expected under-reporting primarily on the less severe UEs and on 'busy days'. Since the employee who identifies a UE is also the one responsible for reporting it, RTTs report more UEs than other staff members. This is partly a consequence of the large number of tasks they perform throughout the treatment course. It is also related to the numerous procedures that take place before RTTs encounter the patient for treatment. At each of these stages, there is a potential for a UE to occur [13].

Literature can be found on accidental exposures and overall design of an Incident Learning System [14–17]. In a busy clinic an easy and fast system for registration of the UEs is essential. To maintain the commitment to continuous reporting it is relevant to have a national system to obtain relevant feedback. Contemporary radiotherapy is more complex than ever with lots of tasks and information going back and forth between different professional groups – based on our work we believe the combination of an updated RPC and PCC highlights the areas of greatest impact for patient safety.

Over time when data mature more, comprehensive analysis will be made regarding treatment quality on a national and institutional level. The SIG plans to ask all members to score the same UEs in order to explore the consistency across centres. To ensure the best learning for patient safety it is highly recommended that all professions are represented in the local UE panels.

To conclude, the updated Danish system of Patient Centered Code and RPC to report and track (systematic) UEs is recommended. Furthermore, it is strongly recommended for each country to align RPC with local workflow to obtain local site-specific insights into areas requiring improvement.

Acknowledgements

NACP 2026 was financially supported by the Acta Oncologica Foundation.

No funding. Supported (not financial) by DSMF.

Bettina Jeanet Godberg, Tine Waldorf, Troels Jungquist Hansen and Karina Søltoft are gratefully acknowledged for their contribution to the development of the updated Danish PCC and RPC system. All former members of the SIG are also acknowledged for their contributions to informed discussions and their support of national collaboration.

Furthermore, the authors would like to acknowledge Laura Patricia Kaplan for her assistance in improving the language of the manuscript.

Disclosure statements

The authors report that there are no competing interests to declare.

Data availability statement

Data can be shared upon reasonable request.

Ethics declarations & trial registry information

No patient-specific data involved.

Authors' contributions

K.L.J., A.R.J., K.L.G., M.B., H.S., S.B.N.B., B.S. and H.S.R. have all taken part in the writing and editing of the manuscript, the data collection and the analysis of the data.

References

[1] dccg.dk. DCCC radiotherapy – the Danish National Research Center [Internet]. c2020–2026 [cited 2026 Mar 3]. Available from: https://www.dccc.dk/english/national-research-centres/dccc-radiotherapy

[2] Shafic J, Barton M, Noble D, Lemer C, Donaldson LJ. An international review of patient safety measures in radiotherapy practice. Radiother Oncol. 2009;92:15–21. https://doi.org/10.1016/j.radonc.2009.03.007

[3] Malicki J, Bly R, Bulot M, Godet J-L, Jahnen A, Krengli M, et al. Patient safety in external beam radiotherapy, results of the ACCIRAD project: current status of proactive risk assessment, reactive analysis of events, and reporting and learning systems in Europe. Radiother Oncol. 2017;123:29–36. https://doi.org/10.1016/j.radonc.2017.02.016

[4] Consolidation Act of the Health Act, Consolidation Act No. 210 of 27 January 2022 [Internet]. Danish. (Bekendtgørelse af sundhedsloven, LBK nr 210 af 27/01/2022.). [cited 2026 Mars 11]. Available from: https://www.retsinformation.dk/eli/lta/2022/210

[5] Towards Safer Radiotherapy | The Royal College of Radiologists [Internet]. [cited 2026 Mars 11]. Available from: https://www.rcr.ac.uk/our-services/all-our-publications/clinical-oncology-publications/towards-safer-radiotherapy/

[6] Crouch K, Adamson L, Beldham-Collins R, Sykes J, Thwaites D. Learning in radiation oncology: 12-month experience with a new incident learning system. J Med Radiat Sci. 2025;72(1):63–73. https://doi.org/10.1002/jmrs.823

[7] Milosevic M, Angers C, Liszewski B, Drudge CS, Marchand E-L, Bissonnette JP, et al The Canadian National System for Incident Reporting in Radiation Treatment (NSIR-RT) taxonomy. Pract Radiat Oncol. 2016;6(5):334–341. https://doi.org/10.1016/j.prro.2016.01.013

[8] Arnold A, Ward I, Gandhidasan S. Incident review in radiation oncology. J Med Imaging Radiat Oncol. 2022;66(2):291–8. https://doi.org/10.1111/1754-9485.13358

[9] Baehr A, Hummel D, Gauer T, Oertel M, Kittel C, Löser A, et al. Risk management patterns in radiation oncology – result of a national survey within the framework of the Patient Safety in German

Radioation Oncology (PaSaGeRO) project. Strahlenther Onkol. 2023;199(4):350–9. https://doi.org/10.1007/s00066-022-01984-5

[10] Smith S, Wallis A, King O, Moretti D, Vial P, Shafiq J, et al. Quality management in radiation therapy: a 15 year review of incident reporting in two integrated cancer centres. Tech Innov Patient Support Radiat Oncol. 2020;14:15–20. https://doi.org/10.1016/j.tipsro.2020.02.001

[11] The Danish Act on Patient Safety in the Health Care System, Act No. 429 of 10 June 2003 [Internet]. Danish. (Lov om patientsikkerhed i sundhedsvæsenet, LOV nr 429 af 10/06/2003). [cited 2026 Mars 11]. Available from: https://www.retsinformation.dk/eli/lta/2003/429

[12] Kumah A. Adverse event reporting and patient safety: the role of a just culture. Front Health Sev. 2025;5:15811516. https://doi.org/10.3389/frhs.2025.1581516

[13] Wiegmann D, Wood L, Cohen TN, Shappell SA. Understanding the 'Swiss Cheese Model' and its application to patient safety. J Patient Saf. 2022;18(2):119–23. https://doi.org/10.1097/PTS.0000000000000810

[14] International Atomic Energy Agency (IAEA). Lessons learned from accidental exposures in radiotherapy. Safety Report Series No. 17. Vienna; 2000.

[15] Eaton DJ, Byrne JP, Cosgrove VP, Thomas SJ. Unintended doses in radiotherapy – over, under and outside? Br J Radiol. 2018;91:20170863. https://doi.org/10.1259/bjr.20170863

[16] Cunningham J, Coffey M, Knöös T, Holmberg O. Radiation Oncology Safety Information System (ROSIS) – profiles of participants and the first 1074 incidents reports. Radiother Oncol. 2010;97(3):601–7. https://doi.org/10.1016/j.radonc.2010.10.023

[17] Prieto Martin C, Kelly C, Paulo G, Akata D, Mille E, Pourel N, et al. Recommandations for reporting and learning from patient-related incidents and near misses in radiotherapy, interventional cardiology, nuclear medicine and interventional and diagnostic radiology [Internet]. Final report of the MARLIN study. Publications Office of the European Union; 2026. [cited 2026 Mars 11]. Available from: https://data.europa.eu/doi/10.2833/7103627

ACTA ONCOLOGICA
2026, VOL. 65, 430–436
https://doi.org/10.2340/1651-226X.2026.45682

NACP 2026

MJS PUBLISHING

ORIGINAL ARTICLE

KaSARi: a national framework for standardized, automated, and predictive radiotherapy in Finland

Jan Seppälä[a], Maria Tengström[a], Henri Korkalainen[a,b], Tuomas Virén[a], Juuso T.J. Honkanen[a], Juha Nikkinen[c], Kaisa Lehtiö[c], Annika Ålgars[d], Jani Keyriläinen[d], Sami Suilamo[d], Heidi Nurmi[d], Tanja Skyttä[e], Eeva Boman[e,f], Tero Vatanen[g], Liisa Sailas[g], Ulla-Mari Arkko[g], Kristiina Vuolukka[h], Tuomas Koivumäki[h], Tiina Metsäharju[a], Timo Voivalin[a] and Janne Heikkilä[a]

[a]Department of Radiotherapy, Kuopio University Hospital, Kuopio, Finland; [b]Department of Technical Physics, University of Eastern Finland, Kuopio, Finland; [c]Department of Radiotherapy, Research Unit of Health Sciences and Technology, Oulu University Hospital, University of Oulu, Oulu, Finland; [d]Department of Oncology and Radiotherapy, Turku University Hospital and University of Turku, Turku, Finland; [e]Department of Oncology, Tampere University Hospital, Wellbeing Services County of Pirkanmaa, Tampere, Finland; [f]Department of Clinical Physiology, Nuclear Medicine and Medical Physics, Tampere University Hospital, Wellbeing Services County of Pirkanmaa,Tampere, Finland; [g]Cancer Centre, North Carelia Central Hospital, Joensuu, Finland; [h]Department of Radiotherapy and Oncology, Wellbeing Services County of Central Finland, Jyväskylä, Finland

ABSTRACT

Background and purpose: National radiotherapy (RT) data infrastructures are emerging to support treatment quality assurance and outcome research. However, prospective nationwide integration of full DICOM-RT data with toxicity and patient-reported outcome measures (PROMs) remains uncommon. KaSARi (Key advances in Standardizing, Automating, and predicting Risks in RT) was established to create national research infrastructure for prospective RT data collection and integration in Finland.

Materials and methods: KaSARi is a prospective, consent-based multicenter cohort infrastructure involving Finnish university and central hospitals. All adult patients receiving RT are eligible. DICOM-RT datasets from treatment planning and oncology information systems are linked to clinician-graded toxicity, PROMs and relevant clinical variables. The infrastructure builds on previously validated multi-institutional DICOM workflows and follows FAIR (Findable, Accessible, Interoperable, Reusable) data principles. Predefined work packages support data harmonization, scalable data processing and enable downstream applications such as AI-based segmentation, dose prediction, and outcome modeling.

Results: Patient recruitment started in May 2025. By February 2026, 221 patients had been enrolled at the coordinating center, initially including breast cancer patients and subsequently expanding to all RT indications. Recruitment at a second center began in November 2025, with 37 patients enrolled. The infrastructure is projected to include approximately 29,000 patients over 15 years. Integration of dosimetric, clinical and PROM data is progressing through a phased national implementation as additional centers join the network.

Interpretation: KaSARi represents a national infrastructure prospectively integrating full DICOM-RT datasets with clinician-graded toxicity and PROMs across all RT indications within a unified research protocol. The study provides a foundation for nationwide outcome modeling, harmonization of RT practices and development and validation of data-driven AI-based methods.

ARTICLE HISTORY
Received 10 March 2026
Accepted 11 April 2026
Published 18 May 2026

KEYWORDS
Radiotherapy; national data collection; artificial intelligence; toxicity; quality of life

Introduction

Radiotherapy (RT) is one of the most data-intensive modalities in oncology, generating high-resolution image data, detailed structure delineations, 3D-dose distributions and RT plans for every treated patient. In principle, these data provide a rich source of information for evaluating treatment quality, studying dose–response relationships, and supporting the development of improved treatment strategies. However, in routine clinical practice, the data is typically stored locally within institutional systems and rarely centralized or systematically combined across treatment centers to facilitate large-scale research, benchmarking, or standardization of clinical practices. Consequently, although RT continues to deliver excellent treatment outcomes for many cancer types, the systematic investigation of outcomes in the era of modern RT remains relatively limited. The linkage between delivered dose, clinical decision-making and individual patient outcomes is therefore inconsistently implemented and remains incomplete [1].

Three structural limitations hinder data-driven optimization of RT. Firstly, target delineation remains variable across clinicians

CONTACT Jan Seppälä jan.seppala@pshyvinvointialue.fi Centre of Oncology/Radiotherapy Department (141850), Kuopio University Hospital, Kelkkailijantie 7, rak. 7, 3.krs, 70200 Kuopio, Finland

and institutions and the prospective linkage between contouring practices and clinical outcomes remains absent in routine clinical practice [2, 3]. Secondly, standardized dose prescriptions do not ensure uniform dose distributions. Substantial inter- and intra-institutional variability in target coverage and organ-at-risk exposure has been consistently demonstrated, even when identical clinical prescriptions are applied [4, 5]. Thirdly, reliable individual-level prediction of toxicity, recurrence and survival remains limited. Current predictive models rarely integrate voxel-level dose data, systemic therapy exposure, comorbidities and patient-reported outcome measures (PROMs) within a unified framework [6, 7].

In Denmark and Sweden, national RT infrastructures (DcmCollab and SKvaRT) have addressed important aspects of RT data harmonization [8, 9]. However, Finland has lacked a prospective framework integrating complete DICOM-RT datasets with structured outcome collection. RT data in Finland have historically been stored locally within institutional firewalls, with limited cross-center integration due to technical heterogeneity, governance structures, and data protection constraints. This decentralized landscape has restricted the development of large-scale datasets necessary to support systematic multi-center research and infrastructure-level standardization.

KaSARi (Key advances in Standardizing, Automating, and predicting Risks in RT) was established to fill this gap as a prospective, consent-based national RT research infrastructure. The primary aim of KaSARi is to design, implement and deploy a standardized framework for prospective, multicenter RT data integration in Finland. Rather than functioning solely as a registry or a single outcomes study, KaSARi is structured as a research protocol–driven platform that integrates comprehensive DICOM-RT datasets with structured clinical data, clinician-graded toxicity, and PROMs collected during routine care from patients treated with modern RT techniques. The infrastructure enables systematic aggregation of imaging, treatment planning, dose, and outcome data within a harmonized framework and supports scalable multicenter analyses. This platform further enables downstream applications, including outcome modeling and AI-based method development, as use-cases of the infrastructure rather than primary endpoints of this work. To our knowledge, no previous national initiative has prospectively integrated full DICOM-RT datasets, clinician-graded toxicity and PROMs for all RT indications under a unified study protocol.

Material and methods

In Finland, RT is delivered in 12 cancer centers distributed nationwide, within five university hospital catchment areas. National coordination, harmonization and strategic development of cancer care are conducted through the Finnish Cancer Center (FICAN) network [10]. The Finnish Cancer Registry (FCR), established in 1952, provides population-based incidence and mortality data with national coverage [11, 12]. However, RT-specific information within the FCR and disease-specific registries remains limited and typically includes only treatment intent, start and end dates, and the total prescribed dose. Detailed dose distributions, structure sets and reported toxicity or PROMs are not systematically integrated at the national level.

KaSARi was established to address these limitations by creating a harmonized national infrastructure for standardized RT data integration and research utilization. KaSARi is a prospective, multicenter cohort infrastructure designed to collect standardized RT data from all consenting patients treated with RT at participating Finnish centers. The study functions as a long-term national research infrastructure rather than a single clinical trial, enabling multiple sub-studies within defined research work packages. This manuscript focuses on the design, implementation and early-phase deployment of the infrastructure, rather than reporting clinical or predictive outcomes.

All adult patients receiving RT at participating centers are eligible for inclusion upon provision of written informed consent. No diagnosis-specific exclusion criteria are applied. With incremental inclusion of centers, approximately 29,000 patients are projected to be enrolled over 15 years, reflecting nearly all RT-treated diseases in Finland.

Data collection and integration framework

Original DICOM-RT datasets (RT Dose, RT Structure Set, RT Plan and treatment delivery summaries) are extracted from local treatment planning system (TPS)/oncology information system (OIS) environments. Data extraction is performed within institutional secure networks in compliance with Finnish and EU data protection regulations. Once the data are extracted, it is also pseudonymized (Figure 1). Structure nomenclature is harmonized using a standardized naming convention aligned with internationally recognized standards (AAPM TG-263) [13]. In cases where local naming deviates from the standard, structure labels are mapped to a unified nomenclature using predefined translation rules. Automated and semi-automated quality control procedures are implemented to detect inconsistencies in structure definitions, missing, or incomplete data.

The technical foundation of KaSARi builds upon earlier national development work. In 2019, an automated multi-institutional target delineation workflow was implemented across Finnish RT centers and presented at ESTRO 2019 [14]. That system included secure pseudonymization gateways, DICOM transfer via integration platforms, and centralized autosegmentation services operating within hospital firewalls. The successful implementation of this cross-center workflow demonstrated the feasibility of secure DICOM handling and multi-institutional data exchange within the Finnish RT environment. The architectural principles and secure data-transfer mechanisms developed in that project form part of the technical backbone of the current KaSARi infrastructure.

Objectives

The primary objective of KaSARi is to establish a nationwide, standardized research infrastructure and repository for the

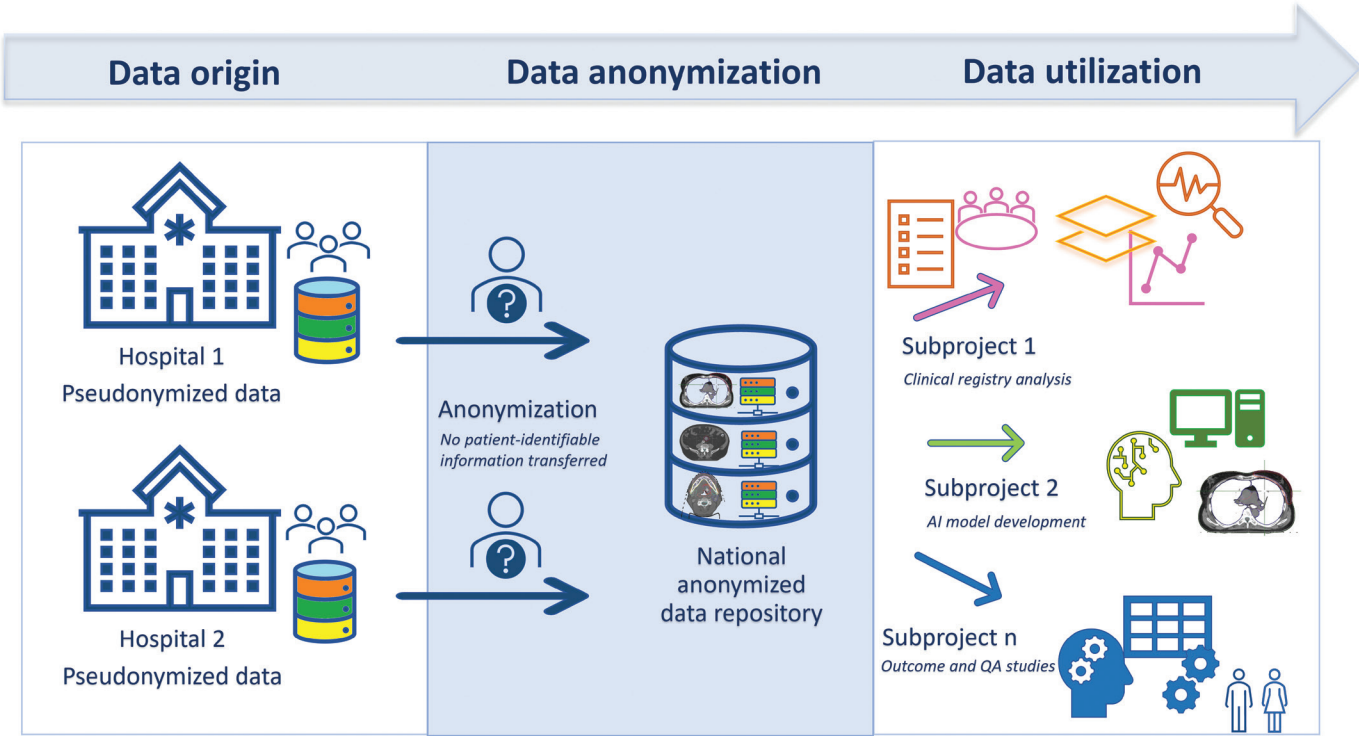


Figure 1. KaSARi data flow from hospital-level pseudonymization to national anonymized repository and research use. QA: quality assurance.

prospective collection and integration of RT imaging, 3D-dose distributions, and clinical outcomes across participating centers (Figure 2). By establishing this unified framework, the infrastructure aims to facilitate large-scale evaluation of treatment delivery, outcomes, and practice patterns in modern RT.

In addition to establishing the infrastructure, KaSARi entails secondary research applications enabled by the platform. These are divided into three research objectives:

1. To develop and validate AI-based segmentation and treatment planning tools (i.e. dose prediction and automated planning) using real-world clinical datasets.
2. To model treatment-related toxicity and quality of life (QoL) outcomes to support data-driven normal tissue complication probability (NTCP) prediction, while accounting for the interactions between RT, systemic therapies, comorbidities, and baseline patient-reported health status.
3. To facilitate federated learning and structured collaboration both nationally and across Nordic RT infrastructures.

Participating centers and governance

The consortium currently includes, Kuopio University Hospital (KUH, coordinating center), North Karelia Central Hospital (PKKS), Oulu University Hospital (OYS), Tampere University Hospital (Tays), Turku University Hospital (Tyks) and University of Eastern Finland (UEF, data science partner). KaSARi operates under a national consortium agreement that defines the common governance, data-sharing principles, and operational framework of the infrastructure. All 12 Finnish municipal RT centers are eligible to join through formal accession to this agreement. This structured yet open model supports nationwide representativeness across diverse RT platforms, clinical workflows and patient populations.

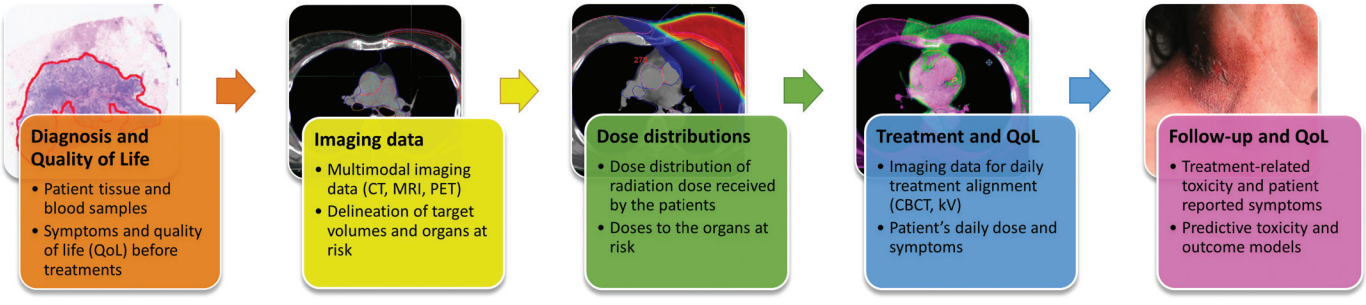


Figure 2. Multimodal data framework of KaSARi, linking baseline clinical variables, imaging and dose data, treatment monitoring, and longitudinal toxicity and quality of life (QoL) outcomes for predictive modeling. CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; CBCT: cone-beam computed tomography; kV: kilovolt imaging.

Results

The research protocol received ethical approval in April 2024 from the Medical Research Ethics Committee of the Wellbeing Services County of North Savo (23/2024). The approval covers the multicenter study framework and allows participating centers to join the study through the national consortium agreement. The following results describe the implementation status and early deployment of the infrastructure, rather than clinical or predictive outcomes.

Patient recruitment started at KUH on May 13, 2025, initially including breast cancer patients only. By the end of 2025, 106 of 206 eligible breast cancer patients had been enrolled, corresponding to a recruitment rate of 51%, demonstrating feasibility of prospective consent-based recruitment within routine clinical workflows, with stable recruitment rates observed during early scaling. Among enrolled patients, 59% completed electronic PROM questionnaires, reflecting early-stage PROM integration within the infrastructure. At the beginning of 2026, recruitment was expanded to include all RT patients at the coordinating center. By the end of February 2026, a total of 221 patients had been enrolled. During the subsequent period from February to March 2026, recruitment continued at a comparable level, with 111 patients enrolled out of 238 RT-treated patients at the coordinating center, corresponding to a recruitment rate of approximately 47% during this early expansion phase. The diagnostic distribution reflects routine clinical practice and includes breast, prostate, lung, and other RT-treated diseases.

Recruitment at Tyks began on November 2025, initially focusing on breast cancer patients. Since initiation, 37 patients have been enrolled, representing approximately 25% of eligible breast cancer patients during the recruitment period. Expansion to additional treatment sites at Tyks is planned. Progressive inclusion of additional Finnish RT centers is ongoing.

Implementation has proceeded in phases to establish recruitment pathways and demonstrate the feasibility of centralized data collection prior to nationwide expansion, reflecting the staged deployment of the infrastructure rather than finalized system maturity. Local data intake is fully operational at the coordinating center, while data intake procedures are being established at participating hospitals. Two additional Finnish RT centers are currently in the process of joining the consortium, and patient recruitment is scheduled to begin shortly in two further hospitals already involved in the network.

The data pipeline has now been successfully established, confirming technical feasibility of standardized multicenter data integration. DICOM-RT datasets are exported from the TPS and OIS to a pseudonymized local intermediate database. Following local validation, the datasets are anonymized and transferred to the national KaSARi repository hosted within a controlled-access research environment at KUH (Figure 1). Metadata, clinical toxicity data, and PROMs are linked via identifiers, enabling longitudinal follow-up across the treatment trajectory, with harmonization procedures applied to ensure consistency of structure definitions, toxicity grading, and PROM datasets across centers.

PROM collection has been successfully integrated at KUH and will be expanded to other participating centers, with current participation rates reflecting early-phase implementation and are expected to improve with workflow standardization and broader system integration. PROMs are collected digitally whenever feasible using validated standardized instruments, including the European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30 [15]. Questionnaires are administered at predefined time points before, during, and after treatment, enabling temporal correlation between delivered dose metrics and patient-reported toxicity. Information on systemic medication, comorbidities, blood sample data, tissue samples, and physician-assessed toxicities is recorded within the OIS and linked to the corresponding DICOM-RT datasets.

Each complete RT dataset, including imaging, dose distributions, and structure sets, averages approximately 0.5–1 GB per patient. With progressive national expansion, annual data growth is projected to reach several terabytes. Scalable and secure storage is implemented within the Finnish national research data infrastructure, ensuring compliance with data protection regulations while enabling controlled research access.

Discussion and conclusion

Despite the increasing availability of detailed digital RT data, systematic infrastructures linking treatment delivery, dose distributions, and clinical outcomes remain limited in many healthcare systems. In particular, large-scale integration of 3D-dose information, structure delineations, and longitudinal PROMs has been inconsistently implemented, restricting opportunities for comprehensive evaluation of treatment practices and outcomes in modern RT. KaSARi was established to address this gap through a national, prospective collection of DICOM-RT data linked with clinician-graded toxicity and PROMs under a unified national governance framework within Finland. Unlike registry-driven infrastructures, KaSARi operates as a research protocol-based architecture, enabling prospective harmonization of technical dose data and clinical outcomes. This work describes the infrastructure and its early implementation, rather than reporting outcome analyses derived from the dataset.

A major strength of KaSARi is its prospective design. Data are collected under a standardized informed consent and harmonized nationally. This prospective approach allows consistent data collection, thereby reducing the risk of missing or incomplete data that often limits retrospective analyses. In addition, prospective governance enables continuous refinement of data collection procedures, allowing new variables, outcome measures, or technical parameters to be incorporated in a coordinated manner across centers as clinical practice evolves. This flexibility facilitates the systematic harmonization of datasets and ensures that the infrastructure remains aligned with emerging research needs and technological developments.

Another major strength of KaSARi is the integration of technical RT data with both clinician-assessed toxicity and

PROMs, although current PROM completeness and consent-based inclusion introduce potential limitations in representativeness that require consideration in downstream analyses. By linking 3D-dose distributions and structure data with longitudinal clinical and patient-reported outcomes, the platform enables detailed investigation of dose–response relationships and supports development of NTCP models that incorporate both clinical toxicity grading and patient-centered outcomes. Finally, methodological development, including AI-based approaches, is embedded within the infrastructure and is presented here as an enabled application of the platform rather than a primary outcome of this study. Activities such as validation of autosegmentation methods, development of dose prediction and automated planning approaches, and dose-based predictive modeling are incorporated as predefined research work rather than retrospective secondary analyses. Embedding these activities within the prospective data collection framework enables systematic evaluation of new methods using harmonized multicenter datasets while maintaining close integration with routine clinical workflows.

The consent-based design and incomplete PROM participation introduce potential selection bias that should be considered in future analyses. Patients providing informed consent may differ systematically from non-participants (age, performance status, comorbidity burden), while PROM completion may depend on symptom severity, patient engagement, or technical accessibility, leading to non-random missingness. These factors may affect representativeness and bias downstream analyses, particularly in predictive modeling and AI development.

The observed recruitment rate (~50%) and PROM completion rate (~60%) reflect an early implementation phase of the infrastructure. Contributing factors include variability in the integration of consent procedures into clinical workflows, patient-related characteristics, and logistical aspects related to the introduction of electronic PROM systems and staff familiarity with study processes. These factors are largely modifiable. Ongoing efforts focus on embedding consent procedures into routine clinical pathways, improving usability and accessibility of digital PROM platforms, and increasing staff training and engagement across centers. The prospective design of KaSARi enables continuous monitoring and iterative optimization of recruitment and data collection processes. Accordingly, future analyses will incorporate appropriate missing data handling and bias-adjustment methods, while participation rates and data completeness are expected to improve as workflows become more standardized and integrated into routine practice.

Several Nordic initiatives have pioneered national RT data harmonization. Recent national audit data from Denmark demonstrate both the increasing complexity of RT practice and substantial variability in treatment indications and dose prescriptions, underscoring the need for standardized, prospective data infrastructures [16]. However, their conceptual frameworks differ from that of KaSARi (Table 1). DcmCollab in Denmark provides storage of raw DICOM datasets primarily derived from clinical trial cohorts. The platform can support integration with toxicity grading and PROM data, particularly within structured study settings. This approach offers maximal flexibility for retrospective dosimetric analyses but is typically based on study-specific or optional data linkage rather than a uniformly applied prospective framework. SKvaRT in Sweden aggregates condensed RT variables for all Swedish patients into a national quality registry. The primary focus is quality assurance and clinical benchmarking and data are intentionally limited to predefined dose-volume descriptors rather than full DICOM datasets. KaSARi complements these models by prospectively integrating complete DICOM-RT datasets with clinician-graded toxicity and longitudinal PROMs within an infrastructure framework designed to support, rather than directly report, downstream analytical and predictive studies. In this sense, KaSARi combines the research-oriented granularity of DcmCollab with the nationwide coverage ambition of SKvaRT, while embedding AI development and predictive modeling as explicit infrastructure objectives. The infrastructure is further designed to support direct use of harmonized multimodal data in predictive modeling and AI development without requiring additional study-specific data integration steps.

Early implementation has highlighted several technical and organizational challenges. Foremost among these is the need for sustainable long-term funding for both research activities and information technology infrastructure. The development and maintenance of secure data storage solutions and robust data pipelines require operational continuity that extends beyond traditional project-based funding models. Secondly, national consensus on structure naming conventions and data dictionaries is critical. As observed in other harmonization efforts, target volumes are more challenging to standardize than organs at risk. While KaSARi adopts internationally recognized nomenclature standards (AAPM TG-263) complemented by ESTRO contouring guidelines, there is ongoing national coordination to maintain semantic consistency as clinical practices evolve, including iterative updates to resolve differences in structure definitions across centers.

A specific and important challenge concerns harmonization of toxicity grading and PROM collection across centers.

Table 1. Comparative overview of the Finnish (KaSARi), Danish (DcmCollab) and Swedish (SKvaRT) national RT data infrastructures.

Feature	KaSARi	DcmCollab	SKvaRT
Data detail	Full DICOM-RT + PROMs	Full raw DICOM-RT	Condensed RT variables
Collection type	Prospective research protocol	Retrospective trial-linked	Automated registry extraction
Outcome linkage	PROMs, survival, CTCAE, biomarkers	Trial-specific endpoints	Survival & limited toxicity
Primary goal	AI & individual risk modeling	Data fidelity & trial storage	Clinical quality monitoring

PROMs: patient-reported outcome measures; CTCAE: Common Terminology Criteria for Adverse Events.

Although KaSARi aims to implement common terminology criteria for adverse events (CTCAE) as the uniform framework for clinician-graded toxicity, local practices vary. To address this, shared national guidelines and training will be implemented to support consistent interpretation and application of CTCAE across centers. KUH utilizes the Kaiku Health (Elekta, Stockholm, Sweden) platform for PROM collection [17], whereas other centers employ locally developed or other electronic health record systems. To reduce heterogeneity, KaSARi prioritizes the use of standardized PROM instruments (e.g. EORTC QLQ-C30) and aims to align data collection timepoints across centers to ensure comparability of outcomes. This heterogeneity can affect data comparability and complicates automated longitudinal analyses. Establishment of standardized electronic questionnaires and interoperable national interfaces would therefore be beneficial to ensure consistent PROM acquisition and seamless database integration. However, the active multicenter collaboration within the project is expected to harmonize the clinical practices and enable consistent data collection even with differing PROM acquisitions. Importantly, harmonization within KaSARi is implemented as an iterative and continuous process rather than a one-time standardization step, with prospective data collection enabling ongoing refinement, validation and feedback across participating centers.

A critical requirement for the long-term sustainability of KaSARi is the establishment of stable funding models that extend beyond traditional project-based research grants. The development and maintenance of national RT data infrastructures involve continuous costs related to data storage, secure computing environments, system maintenance and dedicated personnel. Currently, KaSARi operates through a combination of research funding and institutional support. However, long-term sustainability will likely require integration of infrastructure funding into national research infrastructure frameworks and healthcare system funding structures, supported by coordinated national-level strategies, for example, through the FICAN.

Although KaSARi is developed within the Finnish healthcare system, several core design principles support its transferability to other national contexts. The use of standardized data formats (DICOM-RT) and adherence to FAIR (Findable, Accessible, Interoperable, Reusable) data principles enable interoperability across different technical environments. The architecture, which combines local data storage with centralized or distributed analysis, allows flexibility in implementation depending on national data governance requirements. In addition, the consortium-based model provides a scalable governance structure that can be adapted to healthcare systems with varying degrees of centralization. However, successful implementation in other countries would depend on factors such as national data protection regulations and availability of digital infrastructure. Federated learning approaches may further facilitate international collaboration by enabling model development and validation across institutions without the need to transfer patient-level data.

In the coming years, KaSARi is expected to expand to the majority of Finnish RT centers with the current work representing an initial phase in establishing a scalable national infrastructure. The consortium-based model enables every municipal RT unit to join under a unified governance and technical framework, ensuring nationwide representativeness while maintaining standardized data acquisition. A key priority is the establishment of a permanent national structure to support harmonization of RT practices. In parallel, dashboard-based feedback tools will be implemented to enable real-time benchmarking and support continuous clinical quality improvement across centers. Further development includes the implementation of federated learning pipelines in collaboration with Nordic partners, enabling the secure exchange and validation of AI models without the transfer of raw patient-level data. This approach supports scalable multicenter model development while preserving data protection and institutional autonomy.

KaSARi will also align with international initiatives on RT data standardization and FAIR data principles to ensure interoperability, long-term sustainability, and research reproducibility. As data standards in radiation oncology continue to evolve, the integration of KaSARi variables with structured and harmonized data models will further strengthen consistency, and enable further national and international collaboration.

Conclusions

KaSARi is establishing Finland's first prospective national RT cohort combining complete DICOM-RT datasets with clinician-graded toxicity and PROM values under a unified research model. This manuscript describes the design and early deployment of the infrastructure, which is intended to support future outcome analyses and predictive modeling. By embedding prospective data collection into routine clinical workflow, KaSARi provides a sustainable foundation for nationwide outcome evaluation, AI-driven treatment optimization, and predictive modeling of radiation-induced toxicity. The infrastructure enables the harmonization of RT practices throughout Finland while maintaining the detailed dose data needed for high-resolution research.

Prospective integration of technical RT data with longitudinal clinical outcomes is expected to improve the comparability of treatments, strengthen the development of predictive models and ultimately support more individualized RT with reduced toxicity and improved patient QoL.

Acknowledgments

The authors acknowledge the participating hospitals, Kuopio, Oulu, Tampere, and Turku University Hospitals; North Karelia Central Hospital; and the University of Eastern Finland, for their commitment to the KaSARi consortium.

The authors have received funding from the Emil Aaltonen Foundation and State Research Funding for university-level health research, KUH, Wellbeing Service County of North Savo (Project number 5654256).

The initiative operates within the framework of the Finnish Cancer Center (FICAN) collaboration network.

NACP 2026 was financially supported by the Acta Oncologica Foundation.

Conflict of interest

The authors report that there are no relevant competing interests to declare.

Data availability statement

The datasets generated are part of the KaSARi national radiotherapy research infrastructure and are not publicly available. Access to the data is restricted to participating centers under the KaSARi governance framework and in accordance with informed consent provided by the patients and applicable data protection legislation.

Ethics declarations & trial registry information

This study has the approval from the Medical Research Ethics Committee of the Wellbeing Services County of North Savo (23/2024, 16.4.2024).

Authors' contributions

JS, HK, JH, JTJH and TVi designed the KaSARi research infrastructure and study concept. JS coordinated the study. HK, JH, JTJH and TVi contributed to the development of the technical data integration framework and methodological infrastructure. MT, KV, TVo and TM contributed to the development of patient recruitment procedures and implementation of PROM data collection. MT, JN, KL, AÅ, JK, SS, HN, TS, EB, TVa, UMA, KV, TK and LS contributed to clinical implementation of the study and patient recruitment at the participating centers. JS drafted the manuscript. All authors contributed to manuscript revision and approved the final version. JS takes responsibility for the integrity of the work as a whole.

References

[1] Lustberg T, van Soest J, Jochems A, Deist T, van Wijk Y, Walsh S, et al. Big Data in radiation therapy: challenges and opportunities. Br J Radiol. 2017;90:20160689. https://doi.org/10.1259/bjr.20160689

[2] Guzene L, Beddok A, Nioche C, Modzelewski R, Loiseau C, Salleron J, et al. Assessing interobserver variability in the delineation of structures in radiation oncology: a systematic review. Int J Radiat Oncol Biol Phys. 2023;115:1047–60. https://doi.org/10.1016/j.ijrobp.2022.11.021

[3] Patrick HM, Souhami L, Kildea J. Reduction of inter-observer contouring variability in daily clinical practice through a retrospective, evidence-based intervention. Acta Oncol. 2021;60:229–36. https://doi.org/10.1080/0284186X.2020.1825801

[4] Batumalai V, Jameson MG, Forstner DF, Vial P, Holloway LC. How important is dosimetrist experience for intensity modulated radiation therapy? A comparative analysis of a head and neck case. Pract Radiat Oncol. 2013;3:e99–106. https://doi.org/10.1016/j.prro.2012.06.009

[5] Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: an inter-institutional study of planners and planning systems. Pract Radiat Oncol. 2012;2:296–305. https://doi.org/10.1016/j.prro.2011.11.012

[6] Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues. Int J Radiat Oncol Biol Phys. 2010;76:S3–9. https://doi.org/10.1016/j.ijrobp.2009.09.040

[7] Puri T. Voxel-based dose-toxicity modeling for predicting post-radiotherapy toxicity: a critical perspective. J Clin Med. 2025;14:7248. https://doi.org/10.3390/jcm14207248

[8] Krogh SL, Brink C, Lorenzen EL, Samsøe E, Vogelius IR, Zukauskaite R, et al. A national repository of complete radiotherapy plans: design, results, and experiences. Acta Oncol. 2023;62:1161–8. https://doi.org/10.1080/0284186X.2023.2270143

[9] Olsson CE, Krogh SL, Karlsson M, Eriksen JG, Björk-Eriksson T, Grau C, et al. Danish and Swedish National data collections for cancer – solutions for radiotherapy. Clin Oncol (R Coll Radiol). 2025;37:103657. https://doi.org/10.1016/j.clon.2024.10.009

[10] Finnish Cancer Center (FICAN). [cited 2026 Mar 10]. Available from: https://fican.fi/

[11] Finnish Cancer Registry. [cited 2026 Mar 10]. Available from: https://cancerregistry.fi/information/

[12] Pukkala E, Engholm G, Højsgaard Schmidt LK, Storm H, Khan S, Lambe M, et al. Nordic Cancer Registries – an overview of their procedures and data comparability. Acta Oncol. 2018;57:440–55. https://doi.org/10.1080/0284186X.2017.1407039

[13] Mayo CS, Moran JM, Bosch W, Xiao Y, McNutt T, Popple R, et al. American Association of Physicists in Medicine Task Group 263: standardizing nomenclatures in radiation oncology. Int J Radiat Oncol Biol Phys. 2018;100:1057–66. https://doi.org/10.1016/j.ijrobp.2017.12.013

[14] Heikkilä J, Virsunen H, Voutilainen L, Vuolukka K, Nevantaus A, Haatanen M, et al. Implementing an automated target delineation service in a multi-institutional environment in Finland. Radiother Oncol. 2019;133(Suppl 1):S602. (Abstract PO-1104).

[15] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85:365–76. https://doi.org/10.1093/jnci/85.5.365

[16] Nielsen M, Linaa MB, Hansen VN, Kaplan LP, Lund MD, Nielsen MS, et al. A comprehensive national audit of radiotherapy retreatment numbers, sites and indications. Acta Oncol. 2025;64:872–8. https://doi.org/10.2340/1651-226X.2025.43825

[17] Peltola MK, Lehikoinen JS, Sippola LT, Saarilahti K, Mäkitie AA. A novel digital patient-reported outcome platform for head and neck oncology patients – a pilot study. Clin Med Insights Ear Nose Throat. 2016;9:1–6. https://doi.org/10.4137/CMENT.S40219

ACTA ONCOLOGICA
2026, VOL. 65, 511–517
https://doi.org/10.2340/1651-226X.2026.45688

NACP 2026

MJS PUBLISHING

ORIGINAL ARTICLE

Validation of absolute dose calculation of a single ^{125}I-seed for ophthalmic brachytherapy within distance range 1–15 mm using a diode detector

Anna Rintala[a,b], Vappu Reijonen[a], Assi Valve[a] and Mikko Tenhunen[a]

[a]Helsinki University Hospital, Comprehensive Cancer Center, Helsinki, Finland; [b]Department of Physics, MATRENA, University of Helsinki, Helsinki, Finland

ABSTRACT

Background and purpose: Intraocular tumors can be treated with brachytherapy. Dosimetry of ophthalmic applicators is challenging due to short distance ranges and steep dose gradients. Measurements are sensitive to the structure of the dosimeter, and correction methods specific to ^{125}I-source are needed for validation measurements of dose distribution.

Material and methods: We measured dose profiles of I25.S16 seed using diode detector in a water phantom at 2 and 5 mm above surface of acrylic (PMMA) and gold backing. Volume averaging effect was corrected by the ratio of dose calculated with AAPM TG-43 formalism at the center point and the dose average within the detector element area. The angular sensitivity was measured at a constant radius around the seed. The effect of gold backing was evaluated from measurements with and without gold slab behind the seed. Measurements were normalized to calculated value at the distance of 10 mm from the seed using the air-kerma rate measured with a calibrated well chamber.

Results: 1D/3D gamma comparison pass rates 97% with criteria 0.3 mm/3% and 60% with 0.5 mm/5% were reached for a seed attached to PMMA and gold, respectively, in the range ± 10 mm.

Interpretation: TG-43 calculated single seed distributions in PMMA were equivalent to measurements corrected for angular sensitivity and detector area. Gold-to-water ratio corrected calculations were not similarly comparable, and simple one-dimensional correction method is not accurate enough laterally. Diode dosimetry with these correction methods can be used for quality assurance and validation of clinical dose calculation of ophthalmic applicators.

ARTICLE HISTORY
Received 10 March 2026
Accept 24 April 2026
Published 27 May 2026

KEYWORDS
Iodine-125; diode; dosimetry; brachytherapy

Introduction

Intraocular tumors, such as uveal melanoma, can be treated with ophthalmic brachytherapy. There are many different designs of ophthalmic applicators, in terms of shape and composition of the plaque, radionuclide source employed and how they are affixed to the plaques. One of the most widely used configurations consists of low-energy photon-emitting seeds assembled to the inner surface of a cap made of gold alloy [1, 2].

Dosimetry of brachytherapy sealed sources has been based on the American Association of Physicists in Medicine (AAPM) Task Group 43 (TG-43) dose calculation formalism [3], though there is an aim to transition toward model-based calculation algorithms [4]. Monte Carlo (MC) calculations are the predominant method for determining values for source specific functions used in the TG-43 formalism [5]. One essential limitation of TG-43 is that it assumes full scattering conditions in water [1]. It considers neither tissue heterogeneities nor, more importantly, the backing of ophthalmic applicator. The main effect of the backing is decrease in dose due to the lack of backscattering behind the applicator. It is recommended that for accurate patient dose calculation the effect of backing is taken into account [6]. This is even more important in the case of collimating plaques, where the lips of plaque or single seed slots are used to limit the distribution [2].

As a part of the quality assurance of ophthalmic brachytherapy, calculated dose distribution of an applicator should be validated [1]. However, it is challenging due to the short distance ranges and steep dose gradients from the sources. Small errors in detector positioning translate into large errors in dose deduced from the detector signal. Many types of dosimeters have been used, and detector response artifacts should be minimized or accurately accounted for by appropriate correction coefficients.

Silicon diode detectors have been used for relative dose distribution measurements of ^{125}Iodine (^{125}I) sources, for example, in single seed characterization study [7], investigations of the effect of backing materials [8, 9] and verification of dose distribution of multiple seeds in applicators [10]. The advantages of diodes are small effective size, high current per unit absorbed-dose rate and the possibility of scanning measurements with almost immediate readout in a water phantom system. The

CONTACT Anna Rintala anna.rintala@hus.fi Helsinki University Hospital, Comprehensive Cancer Center, POB 180, FIN00029 HUS, Helsinki, Finland

response of diodes is dependent on photon energy, but for ^{125}I the energy spectrum changes very little with distance [11].

Correction methods verified with single seed measurements are a prerequisite for multiple seed dose distribution measurements because with TG-43 formalism, eye plaque distribution is generated as superposition of single source contributions. Our aim was to develop and validate a measurement set up with a diode detector for dose distribution measurements of custom-made eye plaques [12] used in our clinic. At first, the performance of correction methods for diode signal was demonstrated in water equivalent environment down to very short distances (1 mm) from the ^{125}I-seed center. Finally, the methods were tested against measurements with an applicator surrogate, together with a simple one-dimensional correction for the effect of gold backing applied to calculation.

Material and methods

Measurement setup

We investigated the properties of p-type Si-diode (1110000-1, Sun Nuclear U.S.) for measuring ^{125}I sources. The diode detector has a 0.3 mm entrance window in front of the effective point, and the area of the detector element is a square with 2.3 mm diagonal and thickness of 50 µm.

All diode measurements were carried out in a MP3 water phantom (PTW Freiburg GmbH, Germany) whose detector positioning mechanism supports orthogonal movements with 0.1 mm steps. The coordinate system used in this study is depicted in Figure 1A, where x, y and z represent orthogonal coordinates of the point of interest (measurement) from the seed center and $r = \sqrt{x^2 + y^2 + z^2}$ is the distance from the seed center. α is the angle between x-axis and the measurement position. θ is the angle between z-axis, parallel to the linear source, and the measurement position. Variables r and θ follow the notation used in TG-43 formalism. Custom made PMMA holders (Figure 1B) were attached to the side of the water phantom. The PTW TANDEM electrometer was used for scanning measurements and standalone PTW UNIDOS Romeo electrometer for point measurements with diode bias voltage of 0 V.

The ^{125}I sources that we used in this study are I25.S16 (Eckert & Ziegler Medical, Germany) seeds (physical length 4.5 mm and thickness 0.8 mm, active length 3.5 mm and thickness 0.59 mm). The air-kerma rate of each seed was determined with a calibrated reference-type well chamber (HDR 1000 Plus, single seed holder 70016 (Standard Imaging Inc., U.S.), calibration uncertainty 3.3%) and ambient conditions tracked appropriately [13]). The air-kerma rates of the seeds were between 0.098–0.169 $\mu Gy \cdot m^2 \cdot min^{-1}$ at the time of the measurements.

We calculated dose rates $\dot{D}_{TG-43}(r,\theta)$ with TG-43U1 formalism [3] using radial dose function $g(r)$ fit and anisotropy function $F(r,\theta)$ values given in CLRPv2 database [5, 14]. Consensus value [3] was used for the dose-rate constant in water $\Lambda = 1.012$ cm^{-2} converting air-kerma to absorbed dose at the TG-43 reference point $r_0 = 1$ cm and $\theta_0 = 90°$. The geometry function $G_L(r,\theta)$ was calculated according to the 3.5 mm long linear source.

Area averaging effect

The diode signal averaging effect within the sensitive area is modeled supposing constant sensitivity within the detector area. The dose gradient in the sensitive detector area is dependent on the detector distance r, direction from the source to detector θ and the angular position α of the detector with reference to the source (Figure 1A). The correction for dose averaging within the detector element $C_{avg}(r,\theta,\alpha)$ was evaluated computationally from the ratio of dose rates calculated at the center point (r_c, θ_c) and the average within the

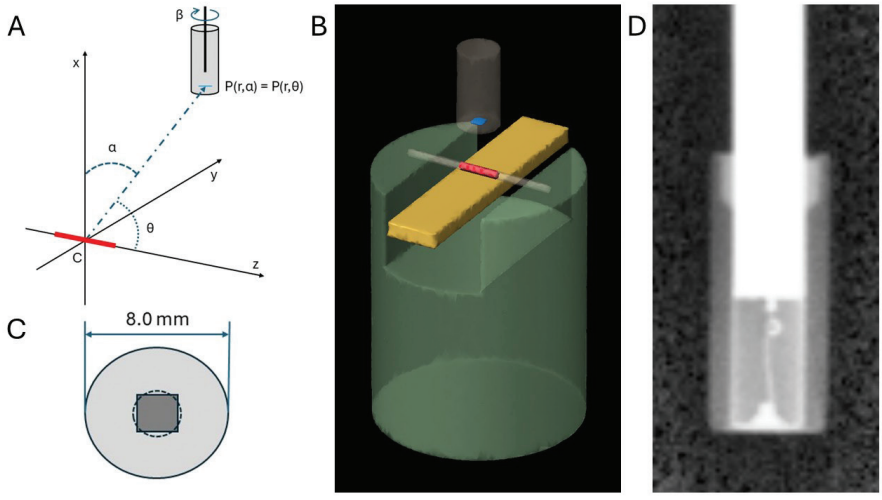


Figure 1. (A) Schematic illustration of the coordinate system of the measurements. The ^{125}I seed (red) lies along the z-axis in the horizontal zy-plane. The angle α denotes the angle between line PC, where point P is the diode position and C is the center of the seed at the origo, and the vertical x-axis. The angle θ of the TG-43 protocol is the angle between the line PC and the z-axis. The angle β is the angle of rotation of the diode detector around its own vertical axis. (B) Simplified illustration of the PMMA holder used in the measurements. The diode detector (gray, effective area in blue) is moved with respect to the seed (red), which is positioned in a plastic catheter which is glued to the holder. The gold slab (yellow) can be positioned behind the seed. (C) The square detector element is approximated as a circle of the same area. (D) X-ray image of the diode detector.

area of detector element approximated as a circle (Figure 1C):

$$C_{avg}(r_c, \theta_c, \alpha_c) = \frac{\dot{D}_{TG-43}(r_c, \theta_c)}{\bar{\dot{D}}_{TG-43}(r, \theta)}. \quad (1)$$

The detector element is square-shaped, but its rotational position within encapsulation is unknown.

Angular sensitivity

Angular sensitivity was evaluated from measurements at a constant radius of 10 mm around the seed axis (θ = 90°) where the dose rate is constant. The seed was positioned in a water filled plastic catheter (wall thickness < 0.3 mm). Charges ΔQ were integrated over time Δt = 1 min at angular range -150°≤ α ≤ 150°, where α is the angle between vertical direction and detector position as in Figure 1A, at 10-degree intervals except 5-degree intervals where the change in the sensitivity is greater. Small corrections were made to the measured values due to the discrete steps of detector positioning mechanism resulting in deviations from the intended radius.

X-ray image of the inner structure of the diode is presented in Figure 1D. Angular sensitivity was measured at four rotational positions (β) around the detector axis at 45-degree intervals. All four curves were normalized to one at angle α = 0° mirrored through it and averaged to get the correction factor $C_{ang}(\alpha)$, which is an inverse of sensitivity values:

$$C_{ang}(\alpha) = \frac{\overline{\Delta Q}(\alpha = 0°) \cdot C_{avg}(r = 10\ \text{mm}, \alpha = 0°)}{\overline{\Delta Q}(\alpha, \beta) \cdot C_{avg}(r = 10\ \text{mm}, \alpha)}. \quad (2)$$

For the correction of profile measurement values, linear interpolation between two nearest measured angular points was used. The angular sensitivity was assumed to be constant as a function of distance r or direction θ from the source.

Effect of gold backing

The effect of the gold slab used in the profile measurements was evaluated from point measurements in the x-direction. Measurements were made with and without the 11 × 44 × 2 mm (width × length × thickness) 24K gold slab just under the seed positioned in a catheter (Figure 1B), and the gold values were divided with water values.

The value of gold-to-water dose ratios $C_{gold}(r)$, where r is the distance from the seed center, used to correct calculated dose values, were evaluated from an experimental fit function

$$C_{gold}(x) = a \cdot \exp(b \cdot x) + c. \quad (3)$$

Thus, the water-to-gold corrected calculated dose rates are:

$$\dot{D}_{TG-43,gold}(r, \theta) = C_{gold}(r) \cdot \dot{D}_{TG-43}(r, \theta). \quad (4)$$

Dose profile measurements

A single ^{125}I seed was attached with silicone glue on the surface and in a slot, respectively, of both acrylic (PMMA) and a gold slab. The depth of the slot was the same as the radius of the seed (0.4 mm). Horizontal centering of the position of the diode detector with respect to the center of the seed was done with profile scans. In the vertical direction, position was determined from the surface of the slab taking the depth of the effective point of the diode into account. Depth doses ("xprof") were measured from 1 to 25 mm above the surface and profiles in both short ("yprof") and long ("zprof") axis directions 2 and 5 mm above the slab surface at range ± 15 mm from the x-axis.

Dose distribution measurements $\Delta Q(r, \theta)$ (integration time Δt = 2 s) were normalized to correspond to dose rate in water for the I25.S16 seed that has air-kerma rate K of 1 mGy min^{-1} at the point $P(r_0 = 10\ \text{mm}, \theta_0 = 0°)$ from the seed center based on the air-kerma rates corrected to the time of the measurement.

The reading of the diode was calibrated to average of corrected point measurements at r_0 = 10 mm in water (seed in a catheter):

$$k_{calib} = \frac{\dot{K}(r_0, \theta_0) \cdot \Lambda}{\overline{\Delta Q}_{norm}(r_0, \theta_0) \cdot C_{avg}(r_0, \alpha = 0)}. \quad (5)$$

The detector signal values were normalized with the calibration coefficient and corrected with the angular sensitivity and area averaging factors:

$$\dot{D}_{meas}(r, \theta) = k_{calib} \cdot C_{ang}(\alpha) \cdot C_{avg}(r, \theta, \alpha) \cdot \Delta Q_{norm}(r, \theta). \quad (6)$$

Measured and calculated dose profiles were compared with 1D/3D gamma (γ) index analysis with distance to agreement/dose difference (DTA/DD) criteria 0.2 mm/2%, 0.3 mm/3% and 0.5 mm/5%. Dose difference was normalized to the calculated value at the calibration point, that is, global normalization was used. The threshold of passing the gamma comparison was set to 90%. The DTA/DD criteria were chosen based on the expected performance limit of the dosimetric method – so that pass rates worse than 100% would indicate the achieved accuracy limit – and not according to the actual clinical needs of the ophthalmic brachytherapy treatment modality. However, considering the size scale of ocular structures, submillimeter criterion is relevant.

Results

Area averaging effect

Examples of area averaging correction factors along or parallel to main axes x, y and z are presented in Figure 2A. In the x-direction, the correction factor is larger at short distances with fall-off close to one at distances larger than 10 mm. In the y- and z-profiles the averaging effect varies from underestimation at short distances from x-axis to overestimation at larger distances. The area correction factor is more even and closer to one overall in profiles with x = 5 mm (not shown). The stepwise changes in the z-direction result from the interpolation of the tabulated anisotropy function $F(r,\theta)$ in the TG-43U1 protocol.

Angular sensitivity

The angular sensitivity of the diode is presented in Figure 2B. Differences are small (on average ≤ 2%pt., at maximum 6.5%pt.) between rotational positions around the diode axis,

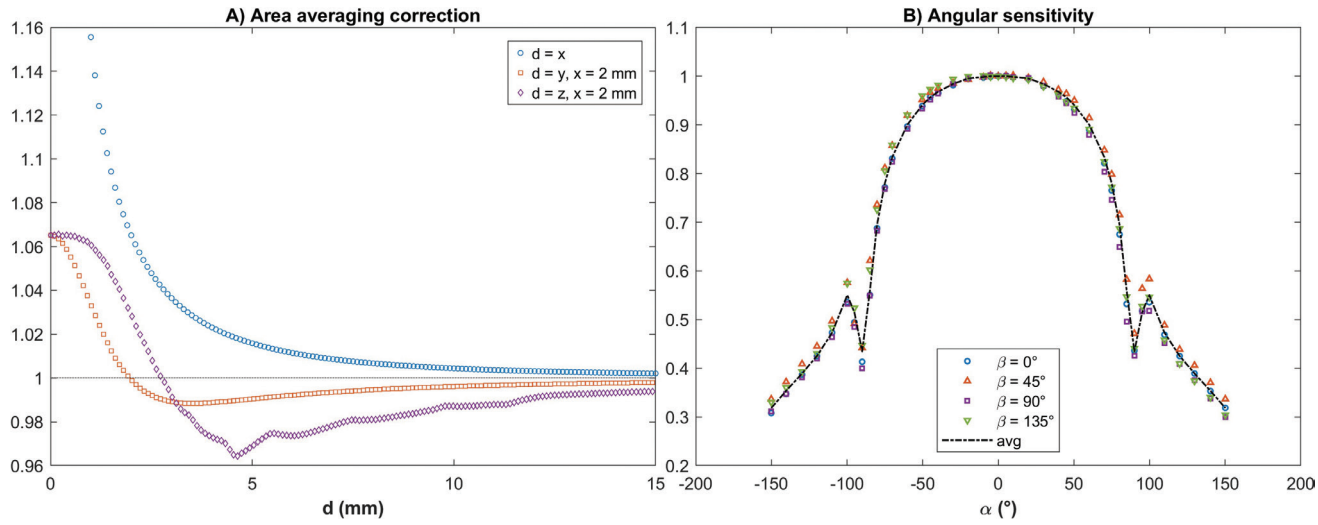


Figure 2. (A) Examples of area averaging within the detector element approximated as a circle along or parallel to the main axes *x*, *y* and *z*. (B) Angular sensitivity of the diode detector measured at four different rotational positions β around the detector's own axis and the averaged sensitivity.

and thus the averaged sensitivity has been used for profile correction. Due to a cylindrical metal filter surrounding the diode effective area to eliminate side scatter, there is an abrupt drop in the sensitivity close to $\alpha = \pm 90°$, and this observation has been used when centering the effective point of the diode detector in *x*-direction in angular sensitivity and gold factor measurements where the reference position cannot be taken from a surface.

Gold correction

The gold-to-water ratio in *x*-direction from the center of the seed is presented in Figure 3. The measured charge values were not corrected for area averaging or angular sensitivity before division since those are equal for both measurements with and without backing. At distances larger than $x = 20$ mm, the ratio reaches a plateau $C_{gold}(x) \approx 0.94$ (not shown).

Dose profiles and gamma index analysis

The calibration coefficient of the diode is $k_{calib} = 3.13 \pm 0.06$ mGy·nC^{-1} (mean ± SD of 10 measurements).

Examples of dose profiles and corresponding γ-values are presented in Figure 4 for measurements with a seed on PMMA-backing and in Figure 5 for a seed on the gold slab. The results of gamma index comparison are presented in Table 1. Profiles were also compared at shorter ranges than the whole measured

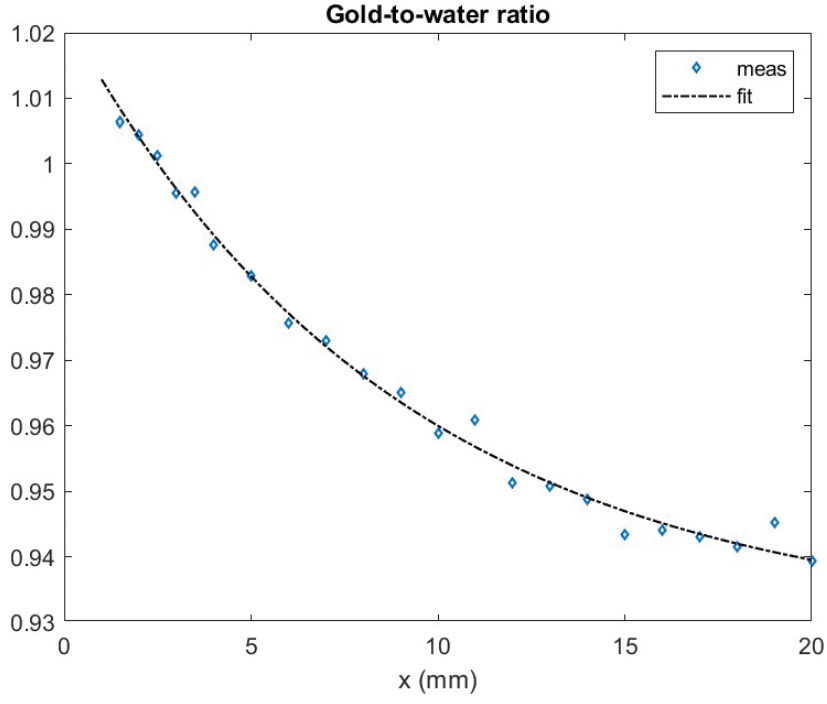


Figure 3. Measured gold-to-water ratio and experimental fit used to correct calculation.

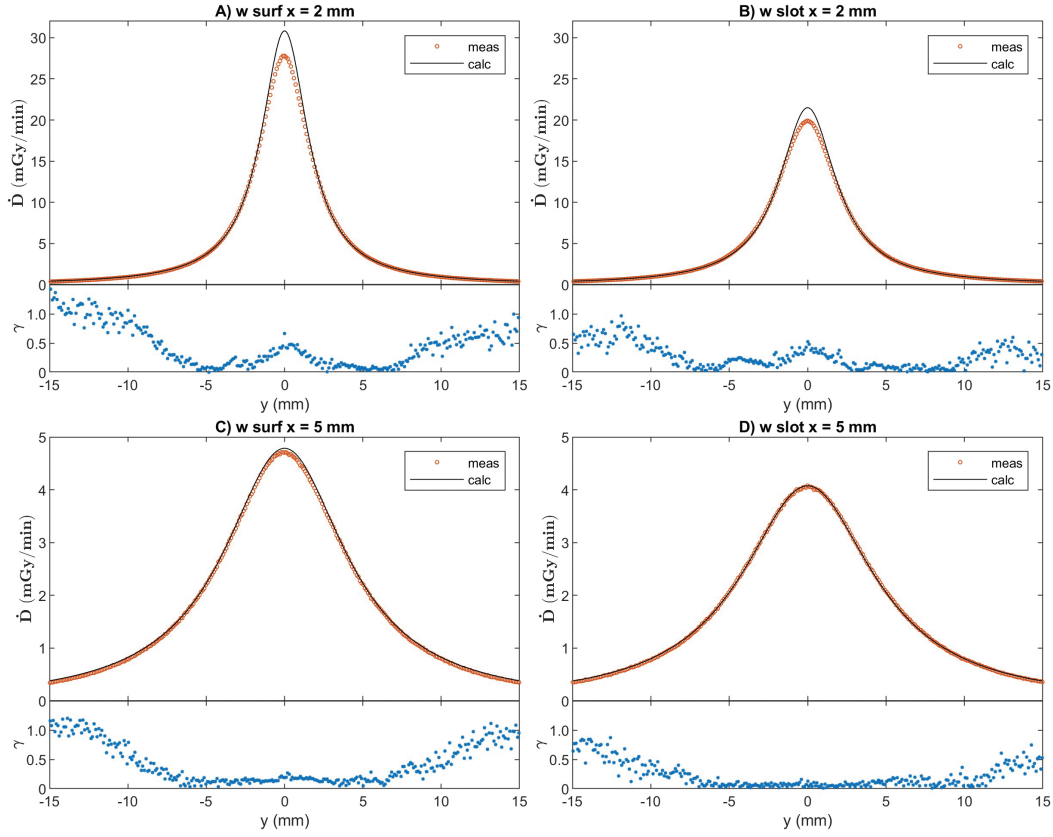


Figure 4. Measured and calculated 1D dose y-profiles with PMMA backing and corresponding γ-values of 1D measurement/3D calculation comparison at 0.3 mm/3% criteria. Abbreviations in figure sections A–D as in Table 1.

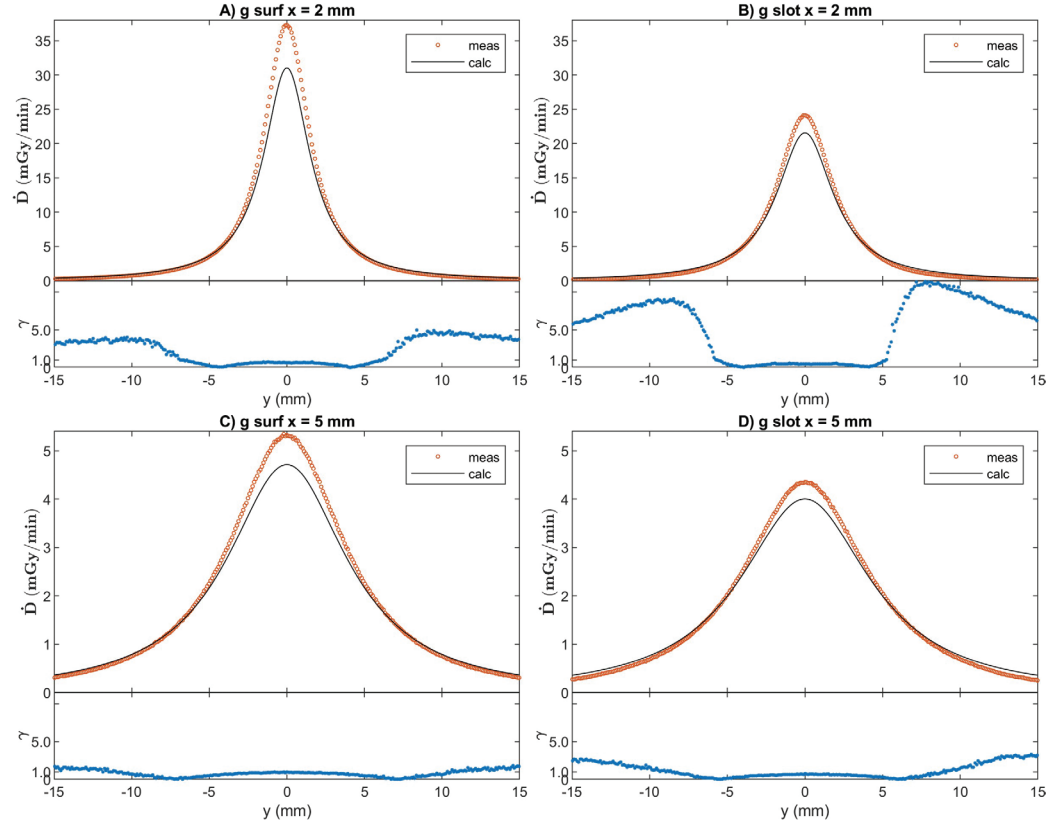


Figure 5. Measured and gold-to-water ratio corrected calculated 1D dose y-profiles with gold backing and corresponding γ-values of 1D measurement/3D calculation comparison at 0.3 mm/3% criteria. Abbreviations in figure sections A–D as in Table 1.

Table 1. The gamma indices, that is, the portion of the points that pass the gamma analysis γ ≤ 1, of the measured dose profiles between indicated distance ranges: for xprof from origo and for yprof and yprof from x-axis.

		0.2 mm/2%					0.3 mm/3%					0.5 mm/5%				
		± 5 mm	± 10 mm	± 15 mm	± 20 mm	± 25 mm	± 5 mm	± 10 mm	± 15 mm	± 20 mm	± 25 mm	± 5 mm	± 10 mm	± 15 mm	± 20 mm	± 25 mm
w	xprof	**95**	**98**	**96**	81	68	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
surf	yprof 2 mm	**99**	**93**	74	-	-	**100**	**99**	**90**	-	-	**100**	**100**	**100**	-	-
	zprof 2 mm	**100**	**90**	72	-	-	**100**	**97**	83	-	-	**100**	**100**	**100**	-	-
	yprof 5 mm	**100**	**99**	72	-	-	**100**	**100**	**91**	-	-	**100**	**100**	**100**	-	-
	zprof 5 mm	**100**	**93**	63	-	-	**100**	**100**	*79*	-	-	**100**	**100**	**100**	-	-
g	xprof	*16*	*7*	*32*	50	61	86	**94**	**96**	**97**	**98**	**100**	**100**	**100**	**100**	**100**
surf	yprof 2 mm	85	50	*33*	-	-	**100**	64	*43*	-	-	**100**	*76*	51	-	-
	zprof 2 mm	*72*	*38*	*26*	-	-	**100**	55	*37*	-	-	**100**	63	*42*	-	-
	yprof 5 mm	*20*	58	*39*	-	-	**100**	**100**	72	-	-	**100**	**100**	**98**	-	-
	zprof 5 mm	*22*	*40*	*26*	-	-	**100**	86	57	-	-	**100**	**98**	*73*	-	-
w	xprof	**95**	**98**	**99**	**99**	**99**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
slot	yprof 2 mm	**100**	**100**	**95**	-	-	**100**	**100**	**100**	-	-	**100**	**100**	**100**	-	-
	zprof 2 mm	**99**	**99**	**92**	-	-	**100**	**100**	**100**	-	-	**100**	**100**	**100**	-	-
	yprof 5 mm	**100**	**100**	**94**	-	-	**100**	**100**	**100**	-	-	**100**	**100**	**100**	-	-
	zprof 5 mm	**100**	**100**	**95**	-	-	**100**	**100**	**100**	-	-	**100**	**100**	**100**	-	-
g	xprof	61	*75*	84	86	84	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
slot	yprof 2 mm	**99**	51	*34*	-	-	**100**	54	*36*	-	-	**100**	60	*40*	-	-
	zprof 2 mm	**100**	55	*37*	-	-	**100**	60	*40*	-	-	**100**	69	*46*	-	-
	yprof 5 mm	80	68	*45*	-	-	**100**	88	59	-	-	**100**	**100**	**72**	-	-
	zprof 5 mm	73	61	*41*	-	-	**100**	82	55	-	-	**100**	**98**	66	-	-

"w" denotes PMMA and "g" gold slab backing, "surf" the surface of the backing and "slot" the 0.4 mm deep (radius of the seed) slot in the backing, "xprof" the depth dose and "yprof" and "zprof" short-axis and long-axis profiles at the denoted distance above the backing. The bolded gamma index values pass the analysis at values over 90% and the values in italics are below 50%.

distance range, and the pass rates are better closer to the *x*-axis and farther from the surface. The calculated dose results with gold backing have overall worse agreement with measurements than those with PMMA backing.

Discussion and conclusion

After applying computational correction for the effective detector size (Eq. 1) and experimentally determined correction for detector angular sensitivity (Eq. 2), we have reached a good accuracy for measuring the dose distribution around a single ^{125}I seed. When calculated values from TG-43 formalism were used as a reference in water equivalent environment, we reached the pass rate better than 97% of tested points with gamma index 0.3 mm/3% and 93% with 0.2 mm/2%, respectively, within the distance range between 1 mm ≤ x ≤ 15 mm and -10 mm ≤ y, z ≤ 10 mm. The gamma index analysis of measurement results in water equivalent environment with reference to TG-43 calculations serves as an approach of uncertainty analysis for future applications of diode measurements in more complex cases.

The positioning of the detector with respect to the seed is fairly easy straight above the seed by measuring two orthogonal (y and z) profiles and centering the detector to the peak. The "depth" direction x transverse to the seed was a more challenging task and two strategies were used: (1) touching the holder surface and resetting x coordinate accordingly with the correction of diode entrance window thickness and (2) measuring Δy = 8 mm shifted x-profile sideways down to below the seed level. This profile presents a drop in the middle, which after angular dependence correction transforms into a peak. Comparing positionings 1 and 2 more than five times, we could confirm the thickness of the diode entrance window indirectly and that the positioning accuracy (as mean difference for coordinate x between methods 1 / 2) and precision (1 SD of the zero position) better than 0.1 mm was reached in all orthogonal directions.

Our dosimeter, the Sun Nuclear scanning diode, is quite old and not commercially available anymore. Nevertheless, we chose it because of its thin entrance window of only 0.3 mm. It is much thinner than the current diodes on market, with most having closer to 1 mm windows. We could mechanically reach the 1 mm distance from the seed center. With this very short distance the detector size forms a significant source of uncertainty, but the approximative theoretical correction appeared to give proper results.

While the volume correction of diode signal is experimental, the area averaging effect is calculated against the theoretical shape of dose distribution from TG-43. While we have not used a fully independent test to prove this theoretical method, the good agreement between the measurements against TG-43 calculation supports the validity of this assumption when the seeds are surrounded in water equivalent medium. At the distance of 10 mm, the volume correction is close to unity. At distances larger than 2.5 mm the volume correction is less than 5%. According to our results, the agreement also seems to be reasonable for seeds embedded in holes of shallow depths (≤ seed cylindrical radius). With tighter collimation (deeper slots) the TG-43 dose profile shape does not apply anymore. The next step is to compare the diode results with measurements using radiochromic film (Gafchromic™, Ashland Inc., U.S.) having minimal area averaging effect.

The effect of backing materials has been studied before [8, 9, 15] and the value of the gold-to-water comparison measurements in our study was the need to calculate accurate enough reference data to compare with our profile measurements. The shape of the curve transverse to the seed is similar to

what has been reported before [8, 9]. One-dimensional path-length corrections based on similar measurements have been used in an ophthalmic treatment planning system [15]. The gamma results are worse for gold backed profiles compared to simple one-dimensional correction as a function of distance from the seed center than in the case of the water equivalent medium. This is likely due to the approximative nature of the applied one-dimensional gold-to-water correction. When the gamma criteria were relaxed to 0.5 mm/5%, the pass rate reached 98% at 5 mm and 60% at 2 mm and for range -10 mm ≤ $y, z \leq 10$ mm. The correction obviously works in the depth direction in which it was determined but fails laterally. Physical modeling of the gold correction would require three-dimensional modeling of the scatter and fluorescence x-rays in the exact seed and gold plaque geometry.

This work forms a basis for further dosimetric studies with ophthalmic applicators loaded with several ^{125}I seeds, part of them embedded in slots for collimation [12]. The correction factors of a single seed as a function of distance and/or direction can be utilized to generate corrections for multiple seed measurements. The correspondence of single seed results should not be interpreted as a validation of clinical dose calculation as a superposition of multiple seed contributions, but the present results validate the measurement methodology. Further measurements with applicators are necessary to evaluate and adjust the performance of dose calculation algorithm used in treatment planning of ophthalmic brachytherapy.

Acknowledgments

The authors would like to thank Mr. Petri Sipilä, M.Sc., goldsmith, for manufacturing the gold slab used in this study.

NACP 2026 was financially supported by the Acta Oncologica Foundation.

This works was supported by the HUS CCC personal grant for AR.

Conflict of interest

AR, AV, MT report no competing interests to declare. VR has received institutional funding for clinical phase-I studies by a sponsor (Bayer) and a dosimetry trial (Pharmtrace).

Data availability statement

Principal author provides data upon request.

Ethics declarations & trial registry information

This research has been conducted in an ethical and responsible manner and is in full compliance with all relevant codes of experimentation and legislation. This study did not involve humans, animals, plants or biological material as research objects. No non-public datasets were used.

Authors' contributions

All authors designed the study. AR and MT conducted the measurements. AR performed the calculations and analysis. AR and MT drafted the manuscript. All authors reviewed, edited and approved the final manuscript.

References

[1] Thomson RM, Furutani KM, Kaulich TW, Mourtada F, Rivard MJ, Soares CG, et al. AAPM recommendations on medical physics practices for ocular plaque brachytherapy: report of task group 221. Med Phys. 2020;47(5):e92–124. https://doi.org/10.1002/mp.13996

[2] Lesperance M, Martinov M, Thomson RM. Monte Carlo dosimetry for 103Pd, 125I, and 131Cs ocular brachytherapy with various plaque models using an eye phantom. Med Phys. 2014;41(3):031706. https://doi.org/10.1118/1.4864474

[3] Rivard MJ, Coursey BM, DeWerd LA, Hanson WF, Huq MS, Ibbott GS, et al. Update of AAPM Task Group No. 43 Report: a revised AAPM protocol for brachytherapy dose calculations. Med Phys. 2004;31(3):633–74. https://doi.org/10.1118/1.1646040

[4] Fletcher EM, Ballester F, Beaulieu L, Morrison H, Poher A, Rivard MJ, et al. Generation and comparison of 3D dosimetric reference datasets for COMS eye plaque brachytherapy using model-based dose calculations. Med Phys. 2024;51(1):694–706. https://doi.org/10.1002/mp.16721

[5] Safigholi H, Chamberland MJP, Taylor REP, Allen CH, Martinov MP, Rogers DWO, et al. Update of the CLRP TG-43 parameter database for low-energy brachytherapy sources. Med Phys. 2020;47(9):4656–69. https://doi.org/10.1002/mp.14249

[6] Chiu-Tsao ST, Astrahan MA, Finger PT, Followill DS, Meigooni AS, Melhus CS, et al. Dosimetry of (125)I and (103)Pd COMS eye plaques for intraocular tumors: report of Task Group 129 by the AAPM and ABS. Med Phys. 2012;39(10):6161–84. https://doi.org/10.1118/1.4749933

[7] Li Z, Fan JJ, Palta JR. Experimental measurements of dosimetric parameters on the transverse axis of a new 125I source. Med Phys. 2000;27(6):1275–80. https://doi.org/10.1118/1.599005

[8] Cygler J, Szanto J, Soubra M, Rogers DW. Effects of gold and silver backings on the dose rate around an 125I seed. Med Phys. 1990;17(2):172–8. https://doi.org/10.1118/1.596588

[9] Meli JA, Motakabbir KA. The effect of lead, gold, and silver backings on dose near 125I seeds. Med Phys. 1993;20(4):1251–6. https://doi.org/10.1118/1.597166

[10] Knutsen S, Hafslund R, Monge OR, Valen H, Muren LP, Rekstad BL, et al. Dosimetric verification of a dedicated 3D treatment planning system for episcleral plaque therapy. Int J Radiat Oncol Biol Phys. 2001;51(4):1159–66. https://doi.org/10.1016/s0360-3016(01)01768-0

[11] International Commission on Radiation Units and Measurements. ICRU Report 72: dosimetry of beta rays and low-energy photons for brachytherapy with sealed sources. J ICRU. 2004;4(2):2–175. https://doi.org/10.1093/jicru/ndh018

[12] Puusaari I, Heikkonen J, Kivelä T. Effect of radiation dose on ocular complications after iodine brachytherapy for large uveal melanoma: empirical data and simulation of collimating plaques. Invest Ophthalmol Vis Sci. 2004;45(10):3425–34. https://doi.org/10.1167/iovs.04-0066

[13] International Atomic Energy Agency (IAEA). Dosimetry in brachytherapy – an international code of practice for secondary standards dosimetry laboratories and hospitals. IAEA TRS 492. Vienna, Austria; 2023.

[14] Safigholi H, Chamberland MJP, Taylor REP, Allen CH, Martinov MP, Rogers DWO, et al. The CLRP TG-43 parameter database for brachytherapy, version 2: E&Z BEBIG GmbH, IsoSeed I25.S06 or .S16. [cited 2026 Feb 24]. Available from: https://physics.carleton.ca/clrp/egs_brachy/seed_database/I125/IsoSeed_125.S06

[15] Astrahan MA. Improved treatment planning for COMS eye plaques. Int J Radiat Oncol Biol Phys. 2005;61(4):1227–42. https://doi.org/10.1016/j.ijrobp.2004.09.062

ACTA ONCOLOGICA
2026, VOL. 65, 543–552
https://doi.org/10.2340/1651-226X.2026.45677

NACP 2026

MJS PUBLISHING

ORIGINAL ARTICLE

Evaluation of 3D-CRT vs. VMAT for radiotherapy of whole breast with simultaneous integrated boost

Gracinda Johansson[a,b], Rafat Kojoj[a], Sewa Surdashi[a], Martin Olin[a], Emil Fredén[a], Pelin Sen[a], Eija Dahl[c] and Camilla Wendt[c]

[a]Medical Physics, Diagnostics and Technology, Södersjukhuset, Stockholm, Sweden; [b]Department of Clinical Science and Education, Karolinska Institutet, Stockholm, Sweden; [c]Department of Oncology, Södersjukhuset, Stockholm, Sweden

ABSTRACT

Background and purpose: This study evaluates volumetric modulated arc therapy (VMAT) as a potential alternative to three-dimensional conformal radiotherapy (3D-CRT) for whole-breast radiotherapy (RT) with simultaneous-integrated boost (SIB).

Patient/material and methods: Ten left-sided breast cancer patients previously treated in our institution (whole breast with SIB) were selected. Clinical plans were generated using tangential field-in-field 3D-CRT technique. VMAT plans were retrospectively created. Dosimetric evaluation was performed according to our clinical guidelines. A robustness evaluation of the 3D-CRT and VMAT plans against simulated anatomical and setup uncertainties was also performed. Differences between techniques were analysed using a two-sided Wilcoxon signed-rank test with a significance level of 0.05.

Results: Both the 3D-CRT and VMAT plans fulfilled the clinical requirements for target volumes and organs at risk. VMAT plans showed a median increase in heart mean dose of 10.8%, compared to 3D-CRT ($p < 0.05$). Heart maximum dose was reduced by up to 56.7% with VMAT ($p < 0.05$), offering a clinically meaningful advantage. For the ipsilateral lung, no significant difference in mean dose was observed, V16Gy was significantly lower and V4Gy was significantly increased with VMAT. VMAT plans improved dose conformity to the boost planning target volume, ($p < 0.05$). Median monitor units (MU) values of 371 MU (359–466 MU) vs 927 MU (752–1018 MU) were obtained with 3D-CRT and VMAT. VMAT remained robust to simulated uncertainties for whole-breast but showed greater sensitivity in the boost volume.

Interpretation: VMAT was a feasible alternative to 3D-CRT for whole-breast RT with SIB, improving boost dose-conformity and reducing maximum heart dose. However, VMAT requires careful dosimetric evaluation and robustness assessment.

ARTICLE HISTORY
Received 9 March 2026
Accepted 29 May 2026
Published 18 June 2026

KEYWORDS
adjuvant radiotherapy; breast cancer; simultaneous integrated boost; volumetric modulated arc therapy; three-dimensional conformal radiotherapy

Introduction

Breast-conserving surgery followed by adjuvant whole-breast radiotherapy (RT) constitutes a cornerstone in the management of early-stage breast cancer, providing excellent local control and survival outcomes [1–3]. The addition of a tumour bed boost has been shown to further reduce the risk of local recurrence, particularly in younger patients and those with adverse pathological features [4–6]. Traditionally, boost irradiation has been delivered sequentially after completion of whole-breast RT [7]. In recent years, the simultaneous integrated boost (SIB) technique has gained increasing clinical interest, as it allows the delivery of differential dose levels to the whole-breast and tumour bed within a single treatment course, thereby reducing overall treatment time and potentially limiting unnecessary radiation exposure to surrounding organs at risk (OARs).

Three-dimensional conformal radiotherapy (3D-CRT) using tangential fields remains the standard technique for whole-breast RT in many institutions, owing to its robustness and favourable sparing of critical structures. The introduction of field-in-field (FiF) techniques has further improved dose homogeneity and reduced high-dose regions within the breast, contributing to improved treatment quality [8]. Nevertheless, when SIB is delivered using 3D-CRT, relatively large volumes of breast tissue may receive the higher boost dose, which has been associated with an increased risk of late normal tissue toxicity, including fibrosis [9]. Furthermore, achieving optimal dose conformity to the boost volume with 3D-CRT can be challenging, particularly in the presence of complex target geometries, potentially resulting in compromises between target coverage and normal tissue sparing.

Modern inverse-planning techniques such as volumetric modulated arc therapy (VMAT) offer the potential to improve plan quality by achieving superior dose conformity and sharper dose gradients around the boost volume [10]. However, the clinical implementation of VMAT in breast RT remains controversial. A major concern is the increased low-dose exposure to

CONTACT Gracinda Johansson gracinda.johansson@ki.se Sjukhusbacken 10, 118 83 Stockholm, Sweden

large volumes of normal tissue, which has been associated with a potential elevated risk of radiation-induced secondary malignancies [11]. The use of deep inspiration breath hold (DIBH) has been shown to significantly reduce cardiac dose by increasing the distance between the heart and the chest wall, thereby improving cardiac sparing across different RT techniques [12–15]. When combined with advanced delivery methods such as VMAT, DIBH may mitigate some of the disadvantages associated with increased low-dose exposure and heart dose.

Despite the increasing availability of VMAT and its demonstrated dosimetric advantages in selected scenarios, its role as an alternative to well-established tangential 3D-CRT for whole-breast RT with SIB remains institution-dependent. Comparative evaluations based on clinically implemented planning strategies are still limited [16, 17], and systematic dosimetric assessments are essential to support evidence-based clinical decision-making. The aim of this study was to evaluate VMAT as a potential alternative to clinical 3D-CRT for whole-breast RT with SIB in left-sided breast cancer patients treated in DIBH.

Patients/material and methods

Patients and structure delineation

Ten left-sided breast cancer patients previously treated with breast conserving surgery followed by whole-breast RT with SIB at Stockholm South Hospital (Stockholm, Sweden) were selected. Patients with boost volumes located centrally or medially in the breast were deemed pertinent for this study, since treatment plans with 3D-CRT result in large volumes of the breast receiving boost dose levels. The prescription dose was 2.67 Gy for the whole-breast and 3.20 Gy for the boost volume, given in 15 fractions to total dose of 40.05 and 48.0 Gy, respectively. Planning computed tomography (CT) images were acquired on a Siemens Somatom Definition Edge (Siemens Healthineers AG, Erlangen, Germany). A Wing Boards™ (CIVCO Radiotherapy, Iowa, USA) was used to immobilise the patient in head-first, supine position with both arms above the head. The CT images consisted of 3-mm axial slices and were acquired in DIBH. DIBH was implemented using surface-guided radiation therapy (SGRT) with the Catalyst system (C-RAD AB, Uppsala, Sweden), which uses chest surface displacement as a surrogate for inspiration level. Patients were initially positioned in free breathing (FB) using a reference surface that was obtained with the Sentinel system (C-RAD AB, Uppsala, Sweden), during the planning-CT acquisition. From this reference surface, an individualised gating reference point was defined on the patient's chest. A stable reference respiratory curve was also obtained. The gating window was set to 3-mm absolute vertical displacement of the gating point, with a DIBH lift of the reference gating point from baseline of typically 10 mm, individualised per patient. Residual variability in inspiration level is acknowledged as a source of geometric uncertainty and is reflected in the planning target volume (PTV) margins applied.

The target volumes were delineated by a radiation oncologist, following our clinical guidelines [18]. The clinical target volume (CTV) of the whole breast (CTV_L_40.05) included the whole breast, while the clinical target volume of the primary tumour (CTVT) for boost volume (CTVT_L_48.0) included the tumour bed and was delineated based on the surgical clips implanted during the breast conserving surgery. Patient characteristics including the volumes of the CTVT_L_48.0 and CTV_L_40.05 as well as the location of the boost volume in the breast are shown in Table 1.

The PTV for the whole breast and the boost volume were created by adding a uniform margin of 7 mm. Both the CTV and the PTV were cropped 5 mm under the body surface to exclude the build-up region of the photon depth-dose distribution in the dose statistics for the target volumes. The OAR were auto-generated by an atlas-based auto-segmentation (ABAS) software (Elekta AB, Stockholm, Sweden) and corrected by a radiation oncologist in Monaco treatment planning system (TPS) (Elekta AB, Stockholm, Sweden). The OARs considered were the heart, both lungs and the contralateral breast.

Treatment planning

Clinical treatment plans were created in Monaco TPS using tangential FiF 3D-CRT technique. Treatment predominantly employed 6 MV photon beams, with 15 MV photons incorporated selectively as low-weight complementary fields when necessary to improve dose homogeneity and target coverage, particularly in patients with larger breast volumes. The plan

Table 1. Patient and target volume characteristics.

Patient#	CTVT_L_48.0 volume (cm^3)	CTV_L_40.05 volume (cm^3)	Volume ratio (%)	Tumour bed location in the breast
1	24.66	383.75	6.43	Central
2	48.82	867.35	5.63	Central
3	25.38	385.62	6.58	Central
4	45.59	836.10	5.45	Medial
5	12.79	254.37	5.03	Medial
6	17.51	547.21	3.20	Central
7	39.11	683.16	5.73	Central
8	21.21	718.41	2.95	Central
9	45.45	479.47	9.48	Central
10	19.38	373.04	5.20	Central

For each patient, the volumes of the boost CTV (CTVT_L_48.0) and the whole-breast CTV (CTV_L_40.05) are reported together with the volume ratio (CTVT_L_48.0 / CTV_L_40.05) and the anatomical location of the tumour bed within the breast. CTV: clinical target volume.

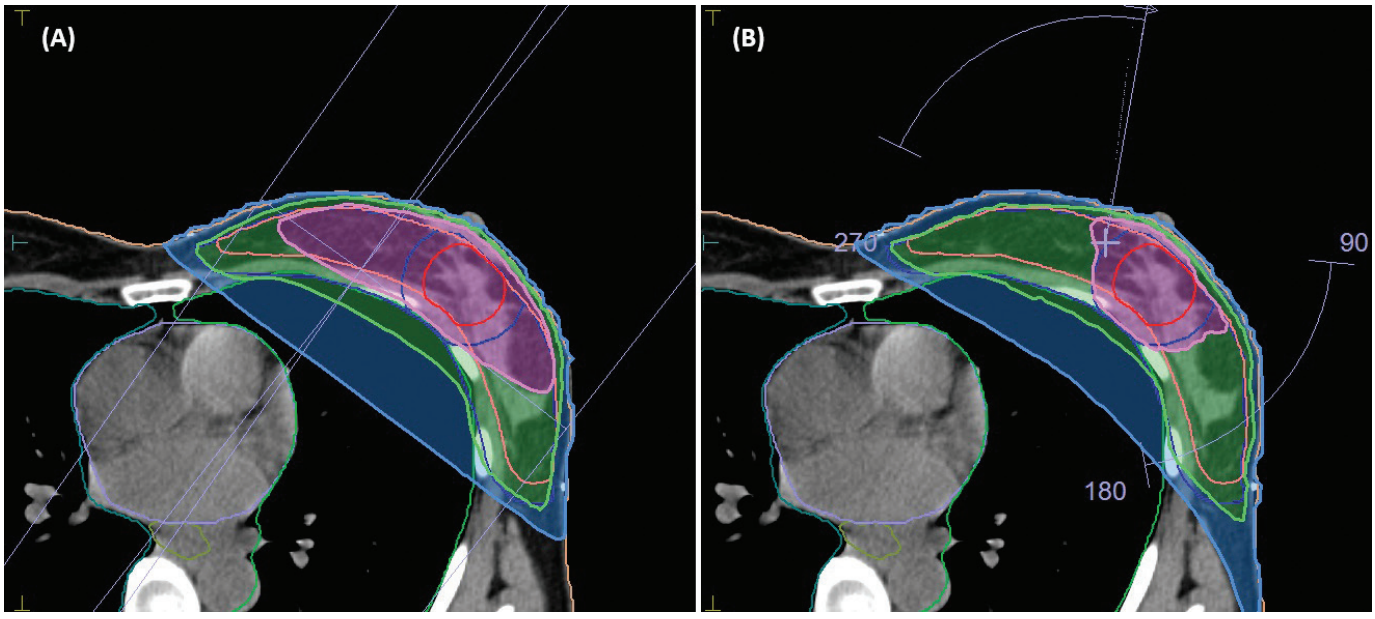


Figure 1. Dose distributions superimposed on a computed tomography (CT) slice shown in transversal view for treatment plans with 3D-CRT (A) and VMAT (B) technique. Dose levels shown are the 95% of the prescribed dose to the boost volume (pink area) and 95% of the prescribed dose to the breast volume (green area). The blue area represents the 20.025 Gy isodose level. The field arrangements used for the 3D-CRT and the VMAT plan are also shown. 3D-CRT: three-dimensional conformal radiotherapy; VMAT: volumetric modulated arc therapy.

isocentre was positioned in the centre of the PTV_L_40.05 longitudinally and vertically. In the lateral direction, the isocentre was placed at ca 8 cm from the midline. The incident beam angles for the medial and lateral tangential fields were optimised to maximise sparing of the heart and the ipsilateral lung. All fields were delivered with a collimator angle of 0°, and the multi-leaf collimator (MLC) was used to block the OARs. Additionally, the tangential fields were extended beyond the patient surface by approximately 2.5 cm to account for expected anatomical changes, such as breast swelling, as well as residual setup and positioning uncertainties affecting the areas close to the patient surface. Dose calculation of the clinical 3D-CRT plans was performed with the collapsed-cone algorithm (dose-to-medium and dose grid of 3 mm). To ensure comparability between techniques, all 3D-CRT plans were recalculated with Monte Carlo algorithm, consistent with the algorithm applied for VMAT planning (dose reported as dose-to-water, a dose grid of 3 mm and a statistical uncertainty of 1% per calculation).

For each patient, VMAT plans were retrospectively created using two tangential VMAT fields, each with three partial arcs of 75°. Photon beam energy was restricted to 6 MV. A collimator angle of 0° was used for all the VMAT fields. The field configuration for the VMAT plans is illustrated in Figure 1. The start angle of the medial VMAT field was adjusted for the individual patients to avoid the contralateral breast and to block the heart. The isocentre position was similar as for the 3D-CRT plans; however, vertically, the isocentre was placed closer to the patient surface. The dose calculation was made using the Monte Carlo algorithm, considering dose deposition in water, a dose grid of 3 mm and a statistical uncertainty of 1% per calculation. The maximum number of control points per arc was set to 90, and the minimum segment width allowed was 1 cm. A skin flash of 2.5 cm was used to improve the dose-coverage of the superficial parts of the breast in the presence of anatomical variation (e.g. breast swelling). A surface margin of 4 mm was applied to account for dose calculation and delivery limitations in the photon beam build-up region (in the patient surface). VMAT optimisation was made using the Pareto approach available in Monaco TPS, with manually assigned weights of the cost-functions for the target volumes and the OARs. All plans were optimised aiming to achieve a clinical acceptable dose coverage to the target volumes while maintaining dose to the OARs as low as possible, based on our clinical dose–volume (DV) criteria given in Table 2.

Plan evaluation

A comparison between the clinical 3D-CRT plans and the VMAT plans was made by means of evaluation of the relevant DV-metrics for the target volumes and the OARs, according to our clinical guidelines (Table 2), which are based on the Swedish national guidelines for breast cancer RT [18]. The conformity index (CI) [19] for the boost volume, defined as the ratio of the volume of the 95% isodose and the volume of the boost Planning target volume of the primary tumor (PTVT)_L_48.0 was also evaluated.

Robustness analysis

The robustness of the 3D-CRT and VMAT plans was evaluated by simulating clinically relevant anatomical and setup uncertainties. Anatomical changes due to breast swelling were modelled by recalculating the original treatment plans on modified CT datasets in which a uniform virtual bolus of 5, 10 or 15 mm thickness was applied to the external breast contour, without modifying the target volumes or replanning. The bolus was assigned

Table 2. Median values and range (minimum–maximum) of target and OAR dose–volume metrics for the clinical 3D-CRT plans and the corresponding VMAT plans.

Structure	DV-metric	3D-CRT	VMAT
Heart	*D*mean < 4 Gy	1.41 (0.98–2.45)	1.57 (1.24–2.23)
	$D_{0.03cm^3}$ < 51.84 Gy	30.1 (6.13–38.25)	16.16 (6.55–28.36)*
Lung_L	*D*mean < 8 Gy	5.46 (2.81–8.42)	5.31 (3.66–7.75)
	V16Gy < 20%	11.67 (4.51–19.87)	8.14 (2.69–15.58)*
	V4Gy < 55%†	23.44 (3.30–35.31)	38.48 (28.9–50.13)*
Breast_R	*D*mean < 2.5 Gy	0.67 (0.44–0.79)	1.69 (1.31–2.33)*
Lung_R	*D*mean < 2 Gy	0.38 (0.29–0.45)	0.74 (0.66–1.00)*
PTV_L_40.05	D98% > 37.25 Gy	37.74 (37.21–38.64)	37.41 (37.28–38.08)
	V45.6Gy < 40%	26.52 (22.46–32.3)	12.04 (7.78–17.42)*
CTV_L_40.05	D98% > 38.05 Gy	38.65 (38.22–39.04)	39.60 (38.74–39.78)
PTVT_L_48.0	D98% > 44.64 Gy	45.49 (43.45–46.18)	45.27 (44.65–46.03)
	CI = V95%/PTV	2.94 (1.69–4.53)	1.23 (1.11–1.43)*
CTVT_L_48.0	*D*mean > 48.0 Gy	48.81 (48.04–49.67)	48.86 (48.68–49.13)
	*D*min > 45.6 Gy	46.54 (45.64–47.60)	46.39 (45.61–47.08)

*P < 0.05 statistically significant; otherwise, not significant. †Taken from the Radiation Therapy Oncology Group (RTOG) 1005 protocol. OAR: organs at risk; 3D-CRT: three-dimensional conformal radiotherapy; VMAT: volumetric modulated arc therapy; DV: dose–volume; CTV: clinical target volume; PTV: planning target volume.

the electron density of water to realistically represent soft tissue swelling. This approach enabled a systematic assessment of the dosimetric impact of progressive breast enlargement on target dose coverage on the original plan, representing a conservative worst-case scenario for dose-coverage degradation. Setup uncertainties were investigated by simulating patient positioning errors through isocentre shifts applied in six directions. A translational displacement of 7 mm was used, corresponding to the clinically used PTV margin. This simulation reflects worst-case residual setup errors after patient positioning and image guidance.

For each simulated scenario, DV metrics were extracted to quantify plan robustness in terms of target dose-coverage. Specifically, D98% for the CTV_L_40.05 and minimum and mean dose (D_{min} and D_{mean}) for the CTVT_48.0 were recorded for all generated treatment plans.

Treatment plan verification

Treatment delivery was performed using an Elekta Versa HD (Elekta AB, Stockholm, Sweden) linear accelerator. For both the 3D-CRT and the VMAT plans, the resulting total monitor units (MU) per plan and the total beam-on time were recorded. Both the 3D-CRT and the VMAT plans were verified using the Delta[4] (ScandiDos AB, Uppsala, Sweden) phantom. A gamma index evaluation with a distance-to-agreement of 3 mm and a percentage dose difference of 3% was performed. The lower threshold of gamma passing rate (GPR) was set to 95%, according to our clinical routine.

Statistical analysis

A two-sided Wilcoxon signed-rank test with a significance level of 0.05 was performed to analyse the differences in paired DV metrics obtained with the clinical 3D-CRT and the VMAT plans.

Results

Plan evaluation

Typical dose ditributions obtained with 3D-CRT and VMAT are shown in Figure 1 and the dose-volume histogram (DVH) comparison is presented in Figure 2. Both the 3D-CRT and VMAT plans fulfilled the predefined clinical requirements for target coverage and

Table 3. Total monitor units (MU) per plan, beam-on time and results for patient-specific quality assurance (PSQA) performed for the 3D-CRT and the VMAT plans.

Patient#	MU		Beam-on time (s)		Gamma index 3 mm/3% (%)	
	3D-CRT	VMAT	3D-CRT	VMAT	3D-CRT	VMAT
1	359	932	67.4	176.0	98.9	99.9
2	373	1003	75.7	174.0	97.8	100.0
3	366	902	72.8	177.0	96.9	100.0
4	392	910	81.0	169.0	96.5	100.0
5	359	955	57.1	166.0	99.9	100.0
6	371	988	70.5	172.0	97.6	100.0
7	385	921	71.8	161.0	98.5	100.0
8	446	1018	108.5	155.0	95.4	100.0
9	370	862	69.6	161.0	97.9	98.6
10	359	752	64.0	173.0	99.9	99.0

3D-CRT: three-dimensional conformal radiotherapy; VMAT: volumetric modulated arc therapy; MU: monitor units.

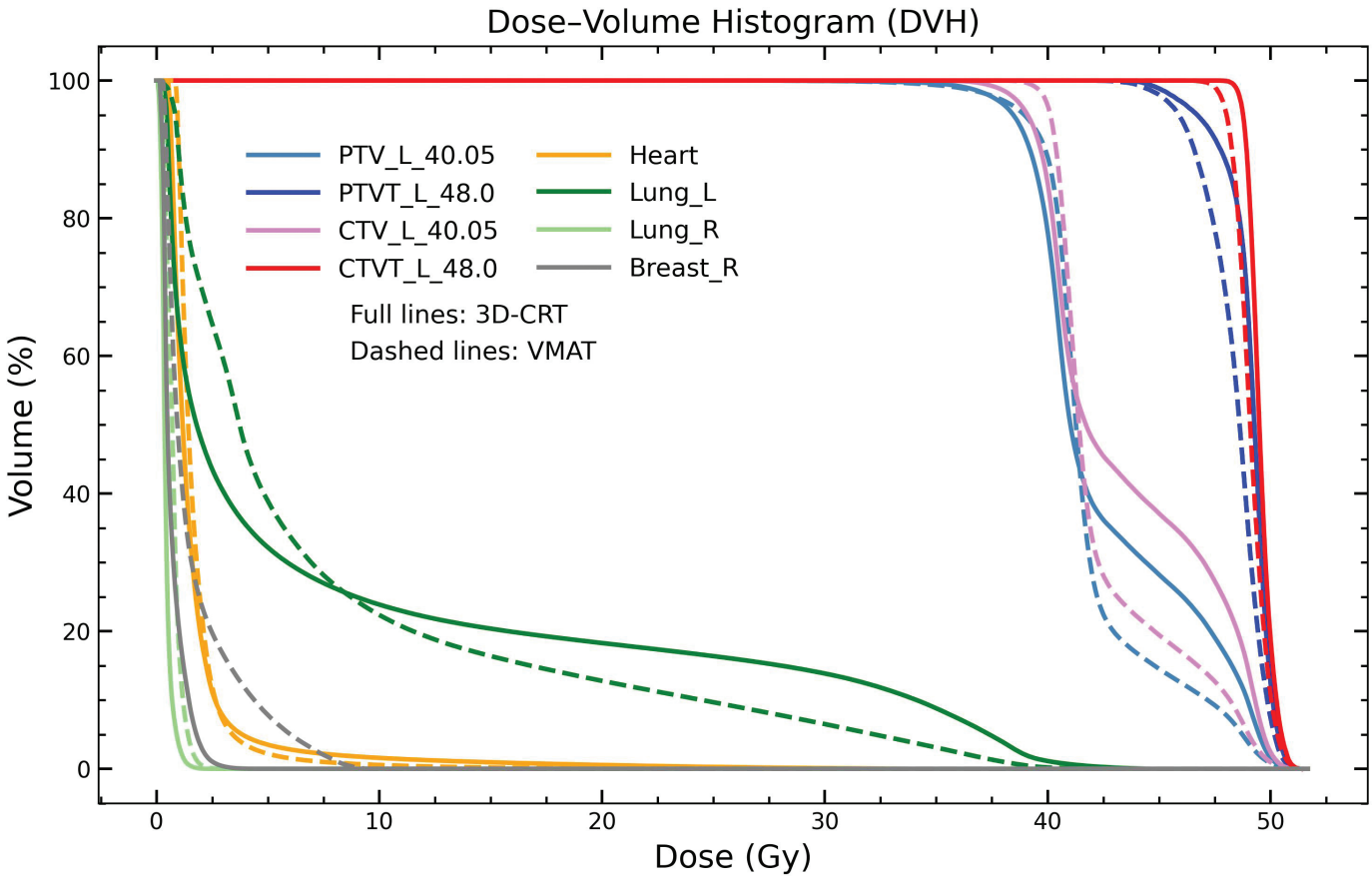


Figure 2. Dose–volume histogram (DVH) comparison between clinical 3D-CRT plan (full lines) and VMAT (dashed lines) for a representative patient. 3D-CRT: three-dimensional conformal radiotherapy; VMAT: volumetric modulated arc therapy; CTV: clinical target volume; PTV: planning target volume.

OAR dose constraints. Compared with 3D-CRT, VMAT was associated with a median increase in mean heart dose of 10.8% (range -9.1 to 26.7%, $p < 0.05$). In contrast, the maximum heart dose was reduced by up to 56.7% with VMAT ($p < 0.05$). For the ipsilateral lung, no significant difference in mean lung dose was observed (median 5.46 Gy with 3D-CRT vs. 5.31 Gy with VMAT, $p > 0.05$), whereas V16Gy was significantly lower with VMAT (8.14%) compared to 3D-CRT (11.67%), $p < 0.05$. The V4Gy of the ipsilateral lung was consistently higher with VMAT (range, 28.9–50.1%) compared to 3D-CRT (range, 3.3–35.3%), $p < 0.05$. For the contralateral breast and the contralateral lung, significantly higher mean dose was registered with VMAT ($p < 0.05$); however, the registered values were below our clinical tolerance levels for these two OARs (Table 2).

VMAT provided significantly improved dose conformity to PTVT_L_48.0 compared with 3D-CRT ($p < 0.05$). The median CI was 1.23 (range 1.11–1.43) for VMAT and 2.94 (range 1.69–4.53) for 3D-CRT.

Robustness evaluation

With the 5-mm simulated swelling, the CTV_L_40.05 D98% criterion (D98% > 38.05 Gy) was fulfilled for seven patients with 3D-CRT and for all patients with VMAT. For CTVT_48.0, the D_{mean} criterion (D_{mean} > 48.0 Gy) was met in 30% (3D-CRT) and 70% (VMAT), while the D_{min} criterion (D_{min} > 45.6 Gy) was fulfilled in 50 and 70%, respectively.

With 10-mm swelling, the CTV_L_40.05 D98% criterion was met in 0% (3D-CRT) and in 90% (VMAT) of the plans. For CTVT_48.0, D_{mean} was fulfilled in 0% (3DCRT) and 10% (VMAT). CTVT_48.0 D_{min} was met in 20% (3DCRT) vs 10% (VMAT).

With the 15-mm swelling, CTV_L_40.05 D98% criterion was not fulfilled in any of the cases with 3D-CRT or VMAT plans. For CTVT_48.0, both D_{mean} and D_{min} criteria were met in one VMAT case and in none of the 3D-CRT plans.

Under simulated ±7-mm isocentre shifts in the three translational directions, the CTV_L_40.05 D98% criterion was met in 83% of all simulated scenarios with 3D-CRT and 93% with VMAT. For CTVT_48.0, the D_{mean} criterion was met in 88 and 87% and the D_{min} criterion was fulfilled in 37 and 15%, for 3D-CRT and VMAT, respectively.

A summary of the robustness evaluation is illustrated in Figures 3, 4 and 5.

Treatment plan verification

Median MU per plan were 371 MU (range 359–466 MU) for 3D-CRT and 927 MU (range 752–1018 MU) for VMAT. Beam-on time ranged from 57 to 109 seconds (0.95–1.82 minutes) for 3D-CRT and from 155 to 177 seconds (2.58–2.95 minutes) for VMAT. Both the VMAT and the 3D-CRT plans met our clinical tolerance values for the PSQA, with a GPR of at least 95.4% (3D-CRT) and 98.6% (VMAT), Table 3.

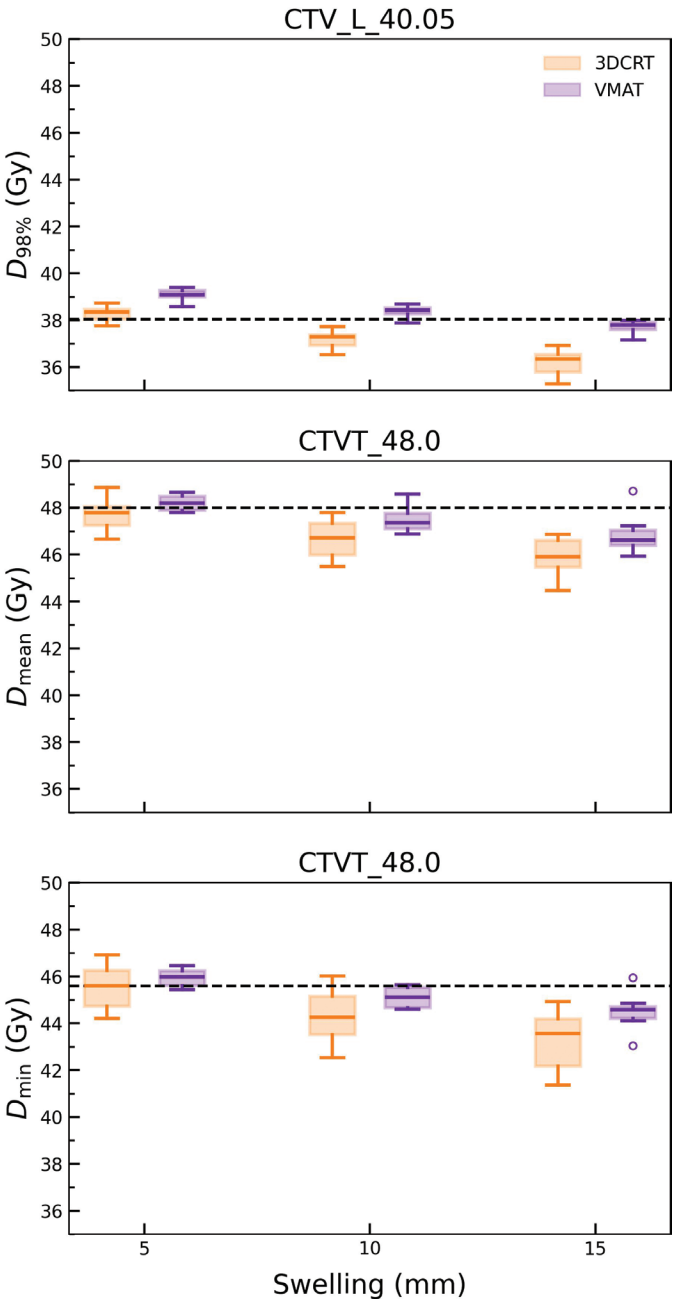


Figure 3. Boxplot showing the dose–volume metric for the CTV_L_40.05 and CTVT_L_48.0 depicting the results of the robustness evaluations of the simulation of breast swelling with uniform thickness of 5, 10 and 15 mm for 3D-CRT (orange) and VMAT (purple). 3D-CRT: three-dimensional conformal radiotherapy; VMAT: volumetric modulated arc therapy; CTV: clinical target volume.

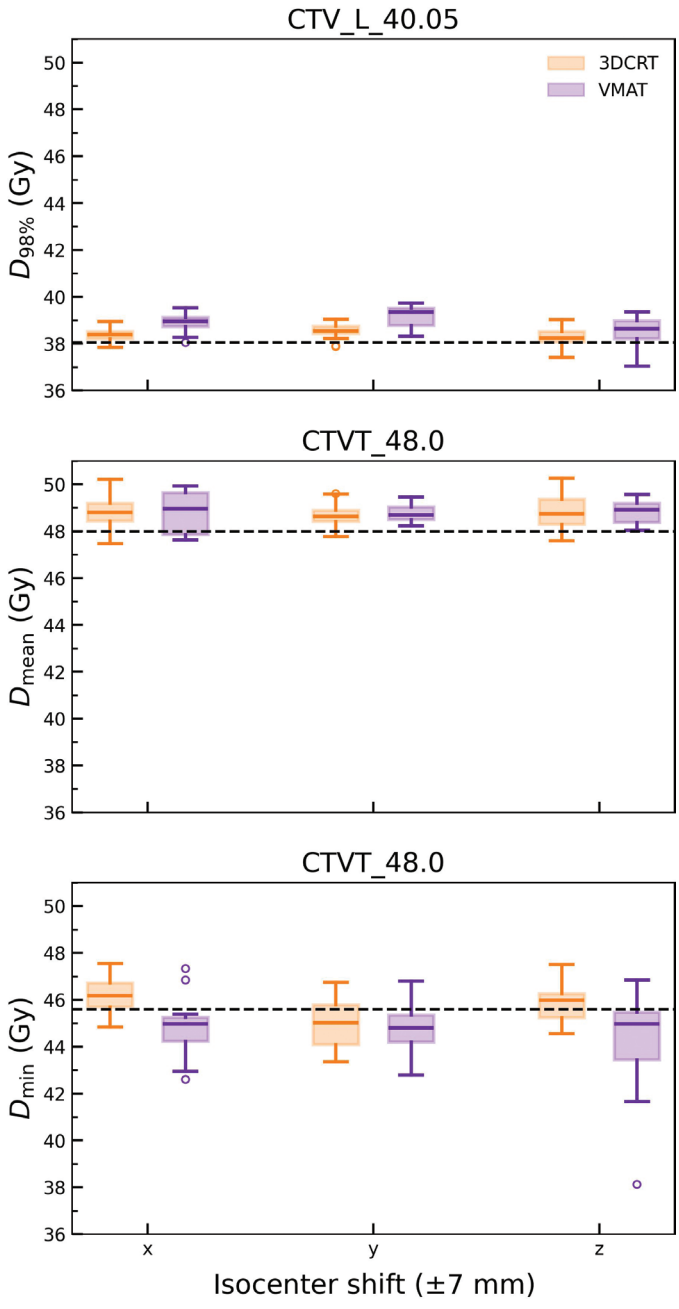


Figure 4. Boxplot illustrating dose–volume metrics for the CTV_L_40.05 and CTVT_L_48.0 depicting the results of the robustness evaluation of the simulated setup errors of 7 mm in three translational axis, for 3D-CRT (orange) and VMAT (purple). 3D-CRT: three-dimensional conformal radiotherapy; VMAT: volumetric modulated arc therapy; CTV: clinical target volume.

Discussion and conclusion

In this study, VMAT was evaluated as a potential alternative to tangential FiF 3D-CRT for left-sided whole-breast RT with SIB in patients treated with DIBH. Both techniques fulfilled institutional planning objectives for target coverage and OAR constraints. However, differences were observed in dose conformity to the boost volume, heart dose, plan robustness and treatment efficiency.

A key finding was the significantly improved dose conformity to the boost PTV achieved with VMAT. This finding is consistent with previous studies demonstrating that inverse-planned techniques generate steeper dose gradients and better conform to complex target geometries [20, 21]. Improved conformity is particularly relevant in SIB treatments, where excessive high-dose exposure of surrounding breast tissue has been associated with increased risk of late toxicity, including fibrosis [4, 5]. The

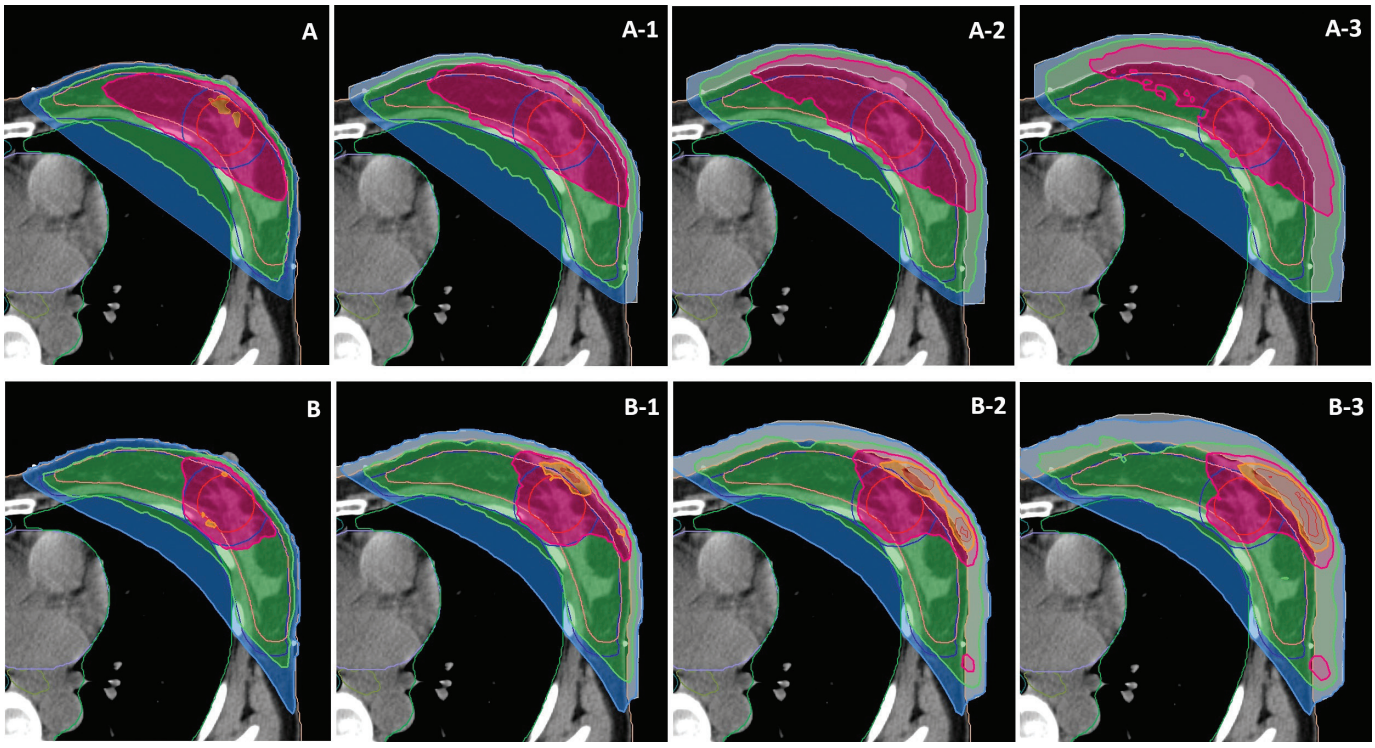


Figure 5. Illustration of dose degradation in the axial view for 3D-CRT (upper panels) and VMAT (lower panels) with the introduction of bolus of 5 mm (A-1, B-1), 10 mm (A-2, B-2) and 15 mm (A-3, B-3). The original plans are shown in A (3D-CRT) and B (VMAT). 3D-CRT: three-dimensional conformal radiotherapy; VMAT: volumetric modulated arc therapy.

European Organisation for Research and Treatment of Cancer (EORTC) boost trial demonstrated higher rates of breast fibrosis in patients receiving boost irradiation, highlighting the importance of limiting unnecessary high-dose exposure [5, 22]. Although clinical outcomes were not assessed in the present study, the reduced high dose spread observed with VMAT may have implications for long-term cosmetic results.

With respect to heart dose, VMAT resulted in a modest but statistically significant increase in mean heart dose compared with 3D-CRT, while significantly reducing maximum heart dose. This reflects the fundamental difference between conventional and intensity-modulated techniques. 3D-CRT limits low-dose spread outside the treatment fields, whereas VMAT redistributes dose more uniformly, reducing focal hotspots at the expense of a broader low-dose bath. Importantly, in the present study, the mean heart dose remained considerably lower than our clinical constraints with both techniques (D_{mean} < 4 Gy). Epidemiologic evidence indicates a relationship between mean heart dose and long-term cardiac risk [13, 23]. The clinical relevance of maximum dose to small cardiac sub-volumes remains less clearly defined [24]. Increasing evidence suggests that mean heart dose alone may not fully capture cardiac risk, as high doses to cardiac substructures such as the left anterior descending artery (LAD) may occur even when global heart doses remain low [25–27]. Wennstig et al. [28] suggested in their retrospective study on breast cancer RT, the inclusion of LAD dose in breast RT planning, with a clear correlation of LAD dose with the risk of radiation-induced stenosis.

The use of DIBH represents an essential component of this treatment approach. DIBH is well established as an effective strategy to reduce heart dose in left-sided breast cancer [12, 13, 29] and likely reduced absolute cardiac exposure for both techniques. The combination of DIBH with conformal techniques such as VMAT may allow improved boost conformity while maintaining acceptable heart sparing, if robustness considerations are addressed.

For the ipsilateral lung, VMAT reduced V16Gy without significantly improving the mean lung dose. The reduction of these DV metrics has also been reported in other studies of VMAT technique for breast cancer RT [17]. Mean lung dose and intermediate DV metrics to the lung (e.g. V15Gy–V20Gy) have been correlated to the risk of radiation-induced pneumonitis following RT [30–33]. Therefore, the relatively lower V16Gy with VMAT could be expected to reduce the incidence of radiation-induced pneumonitis for this patient group. VMAT was also associated with a statistically significant increase in V4Gy for the ipsilateral lung (Table 2), consistent with the low-dose spread inherent to rotational delivery techniques [20, 34]. Nevertheless, all V4Gy values registered for the VMAT plans remained within the constraint used in the Radiation Therapy Oncology Group (RTOG) 1005 study protocol (V4Gy < 55%) [35, 36]. The toxicity implications arising from the low dose lung irradiation were not evaluated in the present study. However, it warrants consideration in patients with pre-existing pulmonary comorbidities, concurrent pulmonary-toxic systemic therapies, or anticipated re-irradiation. When clinically indicated, stricter ipsilateral lung

objectives can effectively reduce V4Gy through constraint reweighting during VMAT optimisation, provided target coverage is not compromised.

Plan robustness represents a central concern when SIB is delivered using highly conformal techniques. Tangential 3D-CRT is traditionally regarded as geometrically robust due to its relatively homogeneous fluence distribution and broader high-dose regions [37]. In contrast, VMAT achieves improved conformity through increased modulation and steeper dose gradients, which may increase sensitivity to geometric perturbations [38, 39]. In the robustness analysis carried out in the present study, CTV_L_40.05 D98% was preserved for both techniques under 5-mm swelling and ±7-mm isocentre shifts. At 10-mm swelling, VMAT maintained dose-coverage to the whole-breast CTV more frequently than 3D-CRT (90% vs 40%), suggesting improved resilience to moderate anatomical expansion. However, at 15-mm swelling, marked deterioration was observed for both techniques (Figure 3). For the boost volume (CTVT_L_48.0), robustness was lower overall. With increasing swelling, both D_{mean} and particularly D_{min} declined for both techniques, without a consistent advantage for VMAT at ≥10 mm (Figure 3). Under simulated ±7-mm shifts (Figure 4), boost D_{min} was fulfilled in 62% of scenarios with 3D-CRT compared with 15% with VMAT. This indicates greater sensitivity of the more conformal VMAT dose distribution to translational setup uncertainties. These findings suggest that improved conformity does not necessarily translate into improved geometric robustness [40], particularly for small high-dose boost volumes.

Treatment efficiency also differed between the evaluated treatment plans (Table 3). While both 3D-CRT and VMAT fulfilled our clinical requirement for PSQA, VMAT required substantially higher MUs and approximately twice the beam-on time compared with 3D-CRT. Beam-on time ranged from 155 to 177 seconds for VMAT and 57–109 seconds for 3D-CRT. In DIBH treatments, longer beam-on time necessitates prolonged or repeated breath-holds, which may increase intra-fraction variability [41, 42]. Although DIBH reproducibility is generally high [29], small inter- and intra-breath-hold variations have been reported [43, 44] and may affect dose delivery, particularly for steep dose gradients. The increased sensitivity of boost D_{min} observed with VMAT may therefore reflect the combined effect of conformal dose distribution and residual geometric variability during repeated breath-holds. Careful breath-hold coaching and motion monitoring are therefore important when VMAT is delivered in DIBH.

This study has a few limitations. The cohort size was limited, and the analysis was restricted to dosimetric parameters, which precludes any conclusions regarding clinical outcomes or treatment-related toxicity. Additionally, only a single VMAT planning strategy was evaluated; alternative arc configurations or hybrid planning approaches may exhibit different robustness characteristics [16, 17, 45, 46]. The analysis reflects routine clinical planning practice and provides relevant information to support institutional decision-making. Furthermore, the robustness analysis for breast swelling was performed by applying a virtual bolus to the planning CT and recalculating dose on the original plan (Figure 5), without extending the CTV or PTV into the swollen volume. While this approach is consistent with published methodology [47, 48] and represents a clinically relevant worst-case scenario, it likely overestimates true coverage in practice.

In conclusion, VMAT fulfilled all predefined planning objectives for whole-breast RT with SIB and provided significantly improved dose-conformity to the boost volume compared with tangential 3D-CRT. VMAT reduced maximum heart dose but was associated with a modest increase in mean heart dose. In the robustness analysis, dose-coverage to whole breast was preserved for both techniques under moderate anatomical changes and setup shifts, whereas the minimum dose given to the boost volume demonstrated greater sensitivity, particularly for VMAT under translational shifts. When combined with DIBH, VMAT represents a feasible alternative to 3D-CRT for selected patients, provided that the trade-off between improved conformity and reduced geometric robustness is carefully considered.

Acknowledgements

NACP 2026 was financially supported by the Acta Oncologica Foundation.

Disclosure statements

The authors report there are no competing interests to declare.

Data availability statement

The data supporting the findings of this study are available from the corresponding author (GJ) upon request.

Ethics declarations & trial registry information

This study used fully anonymised data and was conducted in accordance with institutional guidelines for quality assurance.

Author contributions

GJ, RK, SS, MO, EF, PS, ED and CW designed the study and methodology. GJ prepared the VMAT plans, collected and analysed the data, and wrote the manuscript. RK selected the patients included in this study. SS collected and analysed the data. GJ, RK, SS, MO, EF, PS and CW contributed to data interpretation. MO and PS performed the measurements. All authors reviewed, edited and approved the final version of the manuscript.

References

[1] Darby S, McGale P, Correa C, Taylor C, Arriagada R, Clarke M, et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10 801 women in 17 randomised trials. Lancet.

2011;378:1707–16. https://doi.org/10.1016/S0140-6736(11)61629-2
[2] McGale P, Taylor C, Correa C, Cutter D, Duane F, Ewertz M, et al. Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. Lancet. 2014;383:2127–35. https://doi.org/10.1016/S0140-6736(14)60488-8
[3] Taylor C, Dodwell D, McGale P, Hills RK, Berry R, Bradley R, et al. Radiotherapy to regional nodes in early breast cancer: an individual patient data meta-analysis of 14 324 women in 16 trials. Lancet. 2023;402:1991–2003. https://doi.org/10.1016/S0140-6736(23)01082-6
[4] Bartelink H, Horiot J-C, Poortmans P, Struikmans H, van den Bogaert W, Barillot I, et al. Recurrence rates after treatment of breast cancer with standard radiotherapy with or without additional radiation. N Engl J Med. 2001;345:1378–87. https://doi.org/10.1056/nejmoa010874
[5] Bartelink H, Horiot JC, Poortmans PM, Struikmans H, van den Bogaert W, Fourquet A, et al. Impact of a higher radiation dose on local control and survival in breast-conserving therapy of early breast cancer: 10-year results of the randomized boost versus no boost EORTC 22881-10882 trial. J Clin Oncol. 2007;25:3259–65. https://doi.org/10.1200/JCO.2007.11.4991
[6] Kindts I, Laenen A, Depuydt T, Weltens C. Tumour bed boost radiotherapy for women after breast-conserving surgery. Cochrane Database Syst Rev. 2017;11(11):CD011987. https://doi.org/10.1002/14651858.CD011987.pub2
[7] Coles CE, Haviland JS, Kirby AM, Griffin CL, Sydenham MA, Titley JC, et al. Dose-escalated simultaneous integrated boost radiotherapy in early breast cancer (IMPORT HIGH): a multicentre, phase 3, non-inferiority, open-label, randomised controlled trial. Lancet. 2023;401:2124–37. https://doi.org/10.1016/S0140-6736(23)00619-0
[8] Borghero YO, Salehpour M, McNeese MD, Stovall M, Smith SA, Johnson J, et al. Multileaf field-in-field forward-planned intensity-modulated dose compensation for whole-breast irradiation is associated with reduced contralateral breast dose: a phantom model comparison. Radiother Oncol. 2007;82:324–8. https://doi.org/10.1016/j.radonc.2006.10.011
[9] Bantema-Joppe EJ, Schilstra C, De Bock GH, Dolsma WV, Busz DM, Langendijk JA, et al. Simultaneous integrated boost irradiation after breast-conserving surgery: physician-rated toxicity and cosmetic outcome at 30 months' follow-up. Int J Radiat Oncol Biol Phys. 2012;83(4):e471–7. https://doi.org/10.1016/j.ijrobp.2012.01.050
[10] Teoh M, Clark CH, Wood K, Whitaker S, Nisbet A. Volumetric modulated arc therapy: a review of current literature and clinical use in practice. Br J Radiol. 2011;84:967–96. https://doi.org/10.1259/bjr/22373346
[11] Grantzau T, Overgaard J. Risk of second non-breast cancer among patients treated with and without postoperative radiotherapy for primary breast cancer: a systematic review and meta-analysis of population-based studies including 522,739 patients. Radiother Oncol. 2016;121:402–13. https://doi.org/10.1016/j.radonc.2016.08.017
[12] Swanson T, Grills IS, Ye H, Entwistle A, Teahan M, Letts N, et al. Six-year experience routinely using moderate deep inspiration breath-hold for the reduction of cardiac dose in left-sided breast irradiation for patients with early-stage or locally advanced breast cancer. Am J Clin Oncol Cancer Clin Trials. 2013;36:24–30. https://doi.org/10.1097/COC.0b013e31823fe481
[13] Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Brønnum D, et al. Risk of ischemic heart disease in women after radiotherapy for breast cancer. N Engl J Med. 2013;368:987–98. https://doi.org/10.1056/nejmoa1209825
[14] Kügele M, Mannerberg A, Nørring Bekke S, Alkner S, Berg L, Mahmood F, et al. Surface guided radiotherapy (SGRT) improves breast cancer patient setup accuracy. J Appl Clin Med Phys. 2019;20:61–8. https://doi.org/10.1002/acm2.12700
[15] Freislederer P, Kügele M, Öllers M, Swinnen A, Sauer TO, Bert C, et al. Recent advanced in surface guided radiation therapy. Radiat Oncol. 2020;15:1–11. https://doi.org/10.1186/s13014-020-01629-w
[16] Virén T, Heikkilä J, Myllyoja K, Koskela K, Lahtinen T, Seppälä J. Tangential volumetric modulated arc therapy technique for left-sided breast cancer radiotherapy. Radiat Oncol. 2015;10:79. https://doi.org/10.1186/s13014-015-0392-x
[17] Frengen J, Vikström J, Mjaaland I, Funderud M, Almberg SS, Dybvik KI, et al. Locoregional breast radiotherapy including IMN: optimizing the dose distribution using an automated non-coplanar VMAT-technique. Acta Oncol. 2023;62:1169–77. https://doi.org/10.1080/0284186X.2023.2264488
[18] Regionala cancercentrum (RCC) i samverkan (Regional Cancer Centres in collaboration). 'Nationella vårdprogrammet för bröstcancer' (National Care Program for Breast Cancer) [Internet]. [cited 2026 Jan 25]. Available from: https://cancercentrum.se/samverkan/cancerdiagnoser/brost/kvalitetsregister/
[19] International Commission on Radiation Units and Measurements (ICRU). Prescribing, recording and reporting photon beam therapy. Supplement to ICRU Report 50. ICRU 62. Bethesda (MD): ICRU; 1999.
[20] Nicolini G, Clivio A, Fogliata A, Vanetti E, Cozzi L. Simultaneous integrated boost radiotherapy for bilateral breast: a treatment planning and dosimetric comparison for volumetric modulated arc and fixed field intensity modulated therapy. Radiat Oncol. 2009;4:27. https://doi.org/10.1186/1748-717X-4-27
[21] Fogliata A, Parabicoli S, Paganini L, Reggiori G, Lobefalo F, Cozzi L, et al. Knowledge-based DVH estimation and optimization for breast VMAT plans with and without avoidance sectors. Radiat Oncol. 2022;17(1):200. https://doi.org/10.1186/s13014-022-02172-6
[22] van Parijs H, Reynders T, Heuninckx K, Verellen D, Storme G, De Ridder M. Breast conserving treatment for breast cancer: dosimetric comparison of sequential versus simultaneous integrated photon boost. Biomed Res Int. 2014;2014:827475. https://doi.org/10.1155/2014/827475
[23] Henson KE, McGale P, Taylor C, Darby SC. Radiation-related mortality from heart disease and lung cancer more than 20 years after radiotherapy for breast cancer. Br J Cancer. 2013;108:179–82. https://doi.org/10.1038/bjc.2012.575
[24] Taylor CW, Kirby AM. Cardiac side-effects from breast cancer radiotherapy. Clin Oncol. 2015;27:621–9. https://doi.org/10.1016/j.clon.2015.06.007
[25] Tawfiq N, Guendaoui S, Tantaoui M, Bendahhou K, Hatim G, Chekrine T, et al. Does the mean heart dose remain a valid parameter for assessing early cardiac toxicity following radiotherapy for left-sided breast cancer? Oncol Radioter. 2025;19(9):001–007.
[26] Piroth MD, Baumann R, Budach W, Dunst J, Feyer P, Fietkau R, et al. Heart toxicity from breast cancer radiotherapy: current findings, assessment, and prevention. Strahlentherapie Und Onkol. 2019;195(1):1–12. https://doi.org/10.1007/s00066-018-1378-z
[27] Taylor C, McGale P, Brønnum D, Correa C, Cutter D, Duane FK, et al. Cardiac structure injury after radiotherapy for breast cancer: cross-sectional study with individual patient data. J Clin Oncol. 2018;36:2288–96. https://doi.org/10.1200/JCO.2017.77.6351
[28] Wennstig AK, Garmo H, Isacsson U, Gagliardi G, Rintelä N, Lagerqvist B, et al. The relationship between radiation doses to coronary arteries and location of coronary stenosis requiring intervention in breast cancer survivors. Radiat Oncol. 2019;14(1):40. https://doi.org/10.1186/s13014-019-1242-z
[29] Hayden AJ, Rains M, Tiver K. Deep inspiration breath hold technique reduces heart dose from radiotherapy for left-sided breast cancer. J Med Imaging Radiat Oncol. 2012;56:464–72. https://doi.org/10.1111/j.1754-9485.2012.02405.x
[30] Lind PARM, Wennberg B, Gagliardi G, Fornander T. Pulmonary complications following different radiotherapy techniques for breast cancer, and the association to irradiated lung volume and dose. Breast Cancer Res Treat. 2001;68:199–210. https://doi.org/10.1023/A:1012292019599
[31] Blom Goldman U, Anderson M, Wennberg B, Lind P. Radiation

pneumonitis and pulmonary function with lung dose-volume constraints in breast cancer irradiation. J Radiother Pract. 2014;13:211–17. https://doi.org/10.1017/S1460396913000228

[32] Marks LB, Bentzen SM, Deasy JO, Kong FM, Bradley JD, Vogelius IS, et al. Radiation dose-volume effects in the lung. Int J Radiat Oncol Biol Phys. 2010;76(3 Suppl):S70–6. https://doi.org/10.1016/j.ijrobp.2009.06.091

[33] Karlsen J, Tandstad T, Sowa P, Salvesen Ø, Stenehjem JS, Lundgren S, et al. Pneumonitis and fibrosis after breast cancer radiotherapy: occurrence and treatment-related predictors. Acta Oncol. 2021;60:1651–8. https://doi.org/10.1080/0284186X.2021.1976828

[34] Haciislamoglu E, Colak F, Canyilmaz E, Dirican B, Gurdalli S, Yilmaz AH, et al. Dosimetric comparison of left-sided whole-breast irradiation with 3DCRT, forward-planned IMRT, inverse-planned IMRT, helical tomotherapy, and volumetric arc therapy. Phys Med. 2015;31:360–7. https://doi.org/10.1016/j.ejmp.2015.02.005

[35] Freedman GM, White JR, Arthur DW, Allen Li X, Vicini FA. Accelerated fractionation with a concurrent boost for early stage breast cancer. Radiother Oncol. 2013;106:15–20. https://doi.org/10.1016/j.radonc.2012.12.001

[36] NRG Oncology/Radiation Therapy Oncology Group. ClinicalTrials.gov NCT #: 01349322. RTOG 1005 protocol: a PHASE III trial of accelerated whole breast irradiation with hypofractionation plus concurrent boost versus standard whole breast irradiation plus sequential boost for early-stage breast cancer [Internet]. Amendment 5. 2021. [cited 2026 May 8]. Available from: https://cdn.clinicaltrials.gov/large-docs/22/NCT01349322/Prot_SAP_000.pdf

[37] Krueger EA, Fraass BA, Pierce LJ. Clinical aspects of intensity-modulated radiotherapy in the treatment of breast cancer. Semin Radiat Oncol. 2002;12:250–9. https://doi.org/10.1053/srao.2002.32468

[38] van Mourik A, van Kranen S, Den Hollander S, Sonke JJ, van Herk M, van Vliet-Vroegindeweij C. Effects of setup errors and shape changes on breast radiotherapy. Int J Radiat Oncol Biol Phys. 2011;79:1557–64. https://doi.org/10.1016/j.ijrobp.2010.07.032

[39] Ohri N, Cordeiro PG, Keam J, Ballangrud A, Shi W, Zhang Z, et al. Quantifying the impact of immediate reconstruction in postmastectomy radiation: a large, dose-volume histogram-based analysis. Int J Radiat Oncol Biol Phys. 2012;84(2):e153–9. https://doi.org/10.1016/j.ijrobp.2012.03.026

[40] Unkelbach J, Alber M, Bangert M, Bokrantz R, Chan TCY, Deasy JO, et al. Robust radiotherapy planning. Phys Med Biol. 2018;63(22):22TR02. https://doi.org/10.1088/1361-6560/aae659

[41] Fassi A, Ivaldi GB, de Fatis PT, Liotta M, Meaglia I, Porcu P, et al. Target position reproducibility in left-breast irradiation with deep inspiration breath-hold using multiple optical surface control points. J Appl Clin Med Phys. 2018;19:35–43. https://doi.org/10.1002/acm2.12321

[42] Reitz D, Walter F, Schönecker S, Freislederer P, Pazos M, Niyazi M, et al. Stability and reproducibility of 6013 deep inspiration breath-holds in left-sided breast cancer. Radiat Oncol. 2020;15:121. https://doi.org/10.1186/s13014-020-01572-w

[43] Cerviño LI, Gupta S, Rose MA, Yashar C, Jiang SB. Using suface imaging and visual coaching to improve the repr ducibility and stability of deep-inspiration breath hold for left-breast-cancer radio therapy. Phys Med Biol. 2009;54:6853–65. https://doi.org/10.1088/0031-9155/54/22/007

[44] Alderliesten T, Heemsbergen WD, Betgen A, Topolnjak R, Elkhuizen PHM, van Vliet-Vroegindeweij C, et al. Breast-shape changes during radiation therapy after breast-conserving surgery. Phys Imaging Radiat Oncol. 2018;6:71–6. https://doi.org/10.1016/j.phro.2018.05.006

[45] Voyant C, Pinpin M, Leschi D, Prapant S, Savigny F, Acquaviva MA. Hybrid VMAT-3DCRT as breast cancer treatment improvement tool. Sci Rep. 2023;13(1):23110. https://doi.org/10.1038/s41598-023-50538-x

[46] Byrne M, Archibald-Heeren B, Hu Y, Fong A, Chong L, Teh A. Comparison of semiautomated tangential VMAT with 3DCRT for breast or chest wall and regional nodes. J Appl Clin Med Phys. 2018;19:684–93. https://doi.org/10.1002/acm2.12442

[47] Rossi M, Virén T, Heikkilä J, Seppälä J, Boman E. The robustness of VMAT radiotherapy for breast cancer with tissue deformations. Med Dosim. 2021;46:86–93. https://doi.org/10.1016/j.meddos.2020.09.005

[48] Hennet M, Radonic S, Schneider U, Hartmann M. Retrospective evaluation of a robust hybrid planning technique established for irradiation of breast cancer patients with included mammary internal lymph nodes. Radiat Oncol. 2022;17(1):76. https://doi.org/10.1186/s13014-022-02039-w

ACTA ONCOLOGICA
2026, VOL. 65, 614–623
https://doi.org/10.2340/1651-226X.2026.45698

NACP 2026

MJS PUBLISHING

ORIGINAL ARTICLE

Dose calculation errors for volumetric modulated arc therapy plans of various complexity

Emmanouil Terzidis[a,b], Fredrik Nordström[a,b], Magnus Gustafsson[b], Anna Karlsson[a,b], Julia Götstedt[a,b] and Anna Bäck[a,b]

[a]Department of Medical Radiation Sciences, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [b]Department of Therapeutic Radiation Physics, Biomedical Engineering and Medical Physics, Sahlgrenska University Hospital, Gothenburg, Sweden

ABSTRACT

Background and purpose: Volumetric modulated arc therapy (VMAT) plans often involve small and irregular beam apertures (i.e. high complexity), which can pose challenges for accurate dose calculations. The aim of this study was to investigate plausible dose calculation errors for VMAT plans of various complexity.

Patient/material and methods: Twenty patient cases, each with three VMAT plans of different complexities (i.e. 60 plans in total), were included. All plans were calculated using six different dose calculation methods. It was assumed that greater differences between calculations reflect increased difficulty in accurately estimating dose. Basic performances of the six calculation methods were evaluated based on static fields. Three-dimensional distributions of voxel-wise two standard deviations (2SDs) in percent of local voxel dose were visually evaluated. 2SD volume histograms were analyzed as well as the mean ($2SD_{Mean}$) and maximum 2SD within 2 cm^3 ($2SD_{2cc}$) for different regions of interest.

Results: Higher 2SD values were generally found outside the planning target volume (PTV) compared to inside, particularly in low dose and buildup regions. Average (per treatment site) $2SD_{Mean}/2SD_{2cc}$ values were 0.9–1.2%/1.8–5.1% in the PTV and 1.8–3.1/8.4–18.7% for the region 1 cm outside the PTV. For organs of interest, $2SD_{Mean}/2SD_{2cc}$ values were larger than the equivalent values in the PTV, with highest values up to 5/15% observed for the prostate cases.

Interpretation: Variations between dose calculation methods were larger in organs of interest than in the PTVs. Differences in 2SD distributions between the patient cases were generally larger than differences between the complexity levels.

ARTICLE HISTORY
Received 10 March 2026
Accept 2 July 2026
Published 24 July 2026

KEYWORDS
Dose calculation algorithms; beam model; VMAT; dosimetric uncertainty; plan complexity

Introduction

Accurate and precise dose calculation is crucial for a safe and effective radiotherapy treatment. In photon radiotherapy based on a conventional linear accelerator (LINAC), dose calculation involves two key components. First is the beam model, which provides the beam fluence exiting the treatment head of the LINAC [1]. The second component is the dose calculation algorithm that uses the beam fluence to calculate the dose distribution in a given patient geometry. An accurate beam model and dose calculation algorithm are essential for predicting the dose delivered to the target volume and any nearby organs of interest (OOIs). However, clinical dose calculation methods include approximations that affect the calculated dose distribution in different ways. There are several different types of dose calculation algorithms clinically available, for example, convolution/superposition, linear Boltzmann transport equation solver, and Monte Carlo, each based on different calculation principles. Previous studies have shown that common algorithms generally agree well in homogeneous tissue regions (e.g. [2]). Discrepancies typically arise in heterogeneous regions and in regions where there is a lack of lateral charged particle equilibrium, such as in treatment plans involving small beam openings [3, 4].

Several groups have studied differences in calculated dose distributions between different clinically used dose calculation algorithms, for example, [5–10] often with particular focus on their ability to manage tissue heterogeneities. These studies demonstrate that algorithm performance can diverge considerably in regions with electron density variations. Monte Carlo methods are often regarded as the gold standard in dose calculation due to their ability to accurately simulate particle transport and radiation interactions with matter. However, even the most advanced dose calculation algorithms depend on accurate beam modeling. Treatment plans involving small beam openings or irregular beam aperture shapes (often characterized as complex plans) may require a more detailed beam model to ensure accurate and precise dose calculation [11, 12].

CONTACT Emmanouil Terzidis emmanouil.terzidis@gu.se Department of Medical Radiation Sciences, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, Gula Stråket 2B, 413 46, Gothenburg, Sweden

Supplemental data for this article can be accessed online at https://doi.org/10.2340/1651-226X.2026.45693

To our knowledge, the study from Li et al. [13] is the only one to have specifically examined how disagreement between calculation methods may relate to treatment plan complexity. Their results showed that the investigated complexity metrics had a stronger correlation with differences observed between dose distributions calculated with different calculation methods than with differences between calculated doses and diode-array detector measurements. This indicates that as plan complexity increases, variations between calculation methods become more pronounced.

The aim of this study was to evaluate 3D distributions of plausible errors in the dose calculation for volumetric modulated arc therapy (VMAT) fields of different complexities. For this purpose, the variation between different algorithms and beam models, that is, calculation methods, was studied as a measure to identify regions where accurate dose calculations are challenging.

Patients/material and methods

Beam models and dose calculation algorithms

To highlight regions that are more prone to calculated dose errors, different dose calculation methods with different strengths and weaknesses were used. In this study, it was assumed that larger differences between algorithms indicate greater challenges in accurately estimating the absorbed dose, which, in turn, suggests higher overall dose calculation uncertainty in those regions. Note that the aim of this work was not to determine the absolute accuracy of individual algorithms, but rather to evaluate a range of clinically realistic variations between dose calculation methods and use their in-between differences as an indicator of potential dose calculation uncertainties. This study included six dose calculation methods, implemented in two different treatment planning systems (TPSs), namely, the Eclipse (Siemens Healthineers AG, Erlangen, Germany) and RayStation (RaySearch Laboratories AB, Stockholm, Sweden). The algorithms utilized were as follows:

1. Analytical anisotropic algorithm (AAA) v16.1.0 in Eclipse
2. Analytical anisotropic algorithm (AAA) v18.1.0 in Eclipse
3. Acuros XB (AXB) v16.1.0 in Eclipse
4. Acuros XB (AXB) v18.1.0 in Eclipse
5. Monte Carlo (MC) v.3.0 in RayStation
6. Collapsed cone (CC) v.5.8 in RayStation

All beam model configurations used in this work were created for a 6 MV photon beam from a Varian TrueBeam linac with a Millenium MLC (Siemens Healthineers AG, Erlangen, Germany) and are intended for clinical application. Although not all of them are currently in clinical use. Besides MLC-specific measurements, the beam configurations in Eclipse (i.e. both versions of AAA and AXB algorithms) share the same general commissioning data (e.g. depth doses, profiles, and output factors). AAA version 16.1.0 is the algorithm currently in clinical use. The main difference between version 18.1.0 of AAA and AXB and the earlier 16.1.0 version is an enhanced modeling of the MLC [14] in version 18.1.0. Values of MLC-specific parameter values as well as effective target spot sizes used for the different beam models and algorithms are presented in Table S1a in the supplementary material.

The beam models for the MC and CC algorithms in RayStation were commissioned using a different set of in-house measurement data (measured later in time, but evaluated to be in good agreement, compared to that used for Eclipse configurations) in combination with output factors from Varian reference beam data. Comparisons between Varian reference data and measured data were in good agreement; for example, all depth doses and 96% of lateral profiles were within 95% gamma pass rate using 2%/2 mm criteria. The MLC parameters were acquired in accordance with Saez et al. [15], and primary source width was optimized using profile measurements for a 0.5×0.5 cm^2 beam. Information regarding MLC-specific parameters and primary source widths for the beam models in RayStation can be found in Table S1b in the supplementary material.

For all dose calculation methods, a 1 mm dose grid resolution was utilized. The spacing between CT slices was 2–3 mm. In this study, the focus was placed on calculated dose discrepancies associated with complex treatment plans rather than those arising from heterogeneities within the patient volume. For this reason, the electron and mass density for all voxels in the patient's CT data were set to 1, and the material was set to water. To be able to do a voxel-wise comparison between all dose distributions, some of the dose distributions had to be resampled using trilinear interpolation, to spatially match the AAA v16.1.0 dose distributions. Monte Carlo calculations in RayStation were performed with a statistical uncertainty of 0.1%.

Dose calculations for static fields

Before investigating the variation of different dose calculation methods for VMAT treatment plans, some general aspects of those differences were studied. For this purpose, a number of static fields were created and calculated for 6 MV photons from the Varian TrueBeam linac. The first consisted of a 10 × 10 cm jaw-defined field with the MLC fully retracted and using a source to surface distance (SSD) of 90 cm. This was used to verify that all methods gave the same results in a reference situation. The second field included multi-leaf collimator (MLC) with leaf protrusions as illustrated in Figure 2, using the same jaw positions as the first field. These MLC protrusions were meant to test the performance of the different dose calculation methods specifically when it comes to the MLC field edges and the modeling of the MLC. Differences between dose calculation methods for the two static fields were visually evaluated by extracting lateral dose profiles in the isocenter plane at 10 cm depth. The open field was further evaluated for the different calculation methods by calculating the standard deviation using a coverage factor with $k = 2$ (2SD) of the absorbed dose in the isocenter point. A higher value of 2SD means a larger difference between different

algorithms, which, in this work, is assumed to indicate a higher risk for dose calculation errors. The deviation of the physical field sizes (i.e. both X and Y directions) from the nominal value (i.e. 10 × 10 cm) was also assessed for the different calculation methods. For the static field with MLC protrusions, the 2SD of dose was calculated on a voxel-by-voxel basis across the dose distributions calculated by the six dose calculation methods, resulting in a three-dimensional (3D) distribution of standard deviations. This 3D 2SD map quantified the variance in calculated dose arising from differences between the six dose calculation methods. From this point onward, reference dose refers to the dose calculated by AAA v16.1.0 for each plan since this was the calculation method used in clinical routine. Voxels receiving less than 5% of the reference dose were excluded from this analysis.

The variation between the different dose calculation methods in field size dependence was evaluated by the 2SD of the dose in the isocenter point for MLC-defined fields of decreasing size ranging from 12 × 12 cm down to 1 × 1 cm, with the jaws positioned 2 cm behind the MLC leaf-tips. Additionally, off-axis effects were investigated by positioning the center of a MLC-defined 3 × 3 cm field at lateral offsets of 2, 4, 6, 8, 10, and 12 cm from the isocenter toward the X2 jaw and by evaluating the 2SD of the dose at 10 cm depth in the center of the fields. The jaws were again positioned 2 cm behind the MLC leaf-tips. In all cases, the evaluations were done at 10 cm depth with the SSD set to 90 cm at the central axis. Each field was assigned 240 monitor units (MUs) for the dose calculations, corresponding to 2 Gy at isocenter in the calibration geometry (i.e. 10 × 10 cm field, SSD = 90 cm at 10 cm depth in water).

Dose calculations for treatment plans

In total, 60 VMAT plans were included. The plans were based on CT data and clinical segmentations from 20 cancer patients: five prostate, five head and neck, five lung, and five gynecological cases. The plans were initially optimized and calculated using 6 MV photons with the AAA v16.1.0 dose calculation algorithm (1 mm dose grid). All treatment plans were planned on a Varian TrueBeam linac with the Millenium MLC. Jaw tracking was used for all plans with the exception of prostate cases #1 and #5. Additional information regarding treatment planning parameters for the included VMAT plans can be found in the Supplementary Tables S2–S5.

For each case, a simple and a complex plan were created in addition to the clinical plan, following the methodology described in our previous work [16]. For the simple plans, there was a reduction in MU ranging from 1.3 to 42.6% compared to the respective clinical plan. For the complex plans, there was an increase in MU ranging from 7.4 to 167.4%. Each of the 60 VMAT plans (20 clinical, 20 simple, and 20 complex) was calculated with the six dose calculation methods.

3D 2SD maps were generated for the 60 VMAT plans. Differential 2SD volume histograms were generated from the 3D 2SD distributions for all 60 VMAT plans. These were calculated for the planning target volume (PTV) and a single OOI per treatment site, that is, the rectum for prostate and gynecological cases, the contralateral parotid gland for head and neck cases, and the spinal cord for lung cases. These OOIs were selected as they are considered of high priority for the specific cases that were investigated. Wilcoxon signed-rank tests were performed on pairs of 2SD histogram distributions to determine whether there were significant differences (i.e. $p < 0.05$) in calculated dose variation between plans of different complexity (i.e. clinical vs. simple plans and clinical vs. complex plans). Finally, the maximum 2SD for a volume of 2 cm^3 ($2SD_{2cc}$) and the mean value of 2SD ($2SD_{Mean}$) were calculated inside the PTV, in the region 1 cm outside the PTV and in one OOI per treatment site as described earlier. The region 1 cm outside the PTV was chosen, since it was found to be a region of interest in a previous study, where delivery variations were investigated [16].

Results

Visual evaluation of lateral profiles for a reference situation of a 10 × 10 cm^2 jaw-defined open square field demonstrated consistency across the different calculation methods (Figure 1). The 2SD of the calculated doses in the isocenter was less than 0.5% of the local reference dose, and the field sizes agreed within 0.09 cm of the nominal 10 cm. The highest 2SD for the open part of the field (i.e. within the 8 × 8 cm^2 central part of the field) was 1.8%. This was found in the position at 0.6 and 1.7 cm off axis in the x and y directions, respectively (Figure 1). For the field including MLC protrusions, differences between calculation methods were most pronounced in the MLC-blocked part of the beam near the MLC leaf-tip and in regions of high-dose gradients (Figure 2). For field sizes above 1 × 1 cm^2, the 2SD values for the isocenter dose were below 1.3% of the local reference dose, and it was 2.8% for the 1 × 1 cm^2 field (Figure 3). For off-axis fields, the 2SD values in the central points were below 0.5% at off-axis distances less than 10 cm, reaching up to 1.1% at 12 cm off-axis.

The higher 2SDs were generally observed outside the PTV volume and the highest in the low dose and build up regions (Figure 4). Figure 5 presents the differential 2SD volume histogram curves for PTVs and selected OOIs across the various plan complexities for all cases studied. The histograms demonstrated that the differences between patient cases were generally larger than differences between complexity levels. Statistically significant differences in 2SD histograms between clinical and simple plans were observed in 11 out of 20 cases for PTV and in 7 out of 20 cases for the selected OOIs (Table 1). For comparisons between clinical and complex plans, significant differences were found in 15 out of 20 cases for PTV and 8 out of 20 for OOIs. As shown in Table 2, both $2SD_{Mean}$ and $2SD_{2cc}$ were generally higher for the OOIs compared to the inside of the PTV. For the OOIs, the highest $2SD_{Mean}$ values were observed for the prostate and head & neck cases, while the $2SD_{2cc}$ values for the OOIs were highest for the rectums in the prostate and gynecological cases. The $2SD_{Mean}$ and $2SD_{2cc}$ values inside the

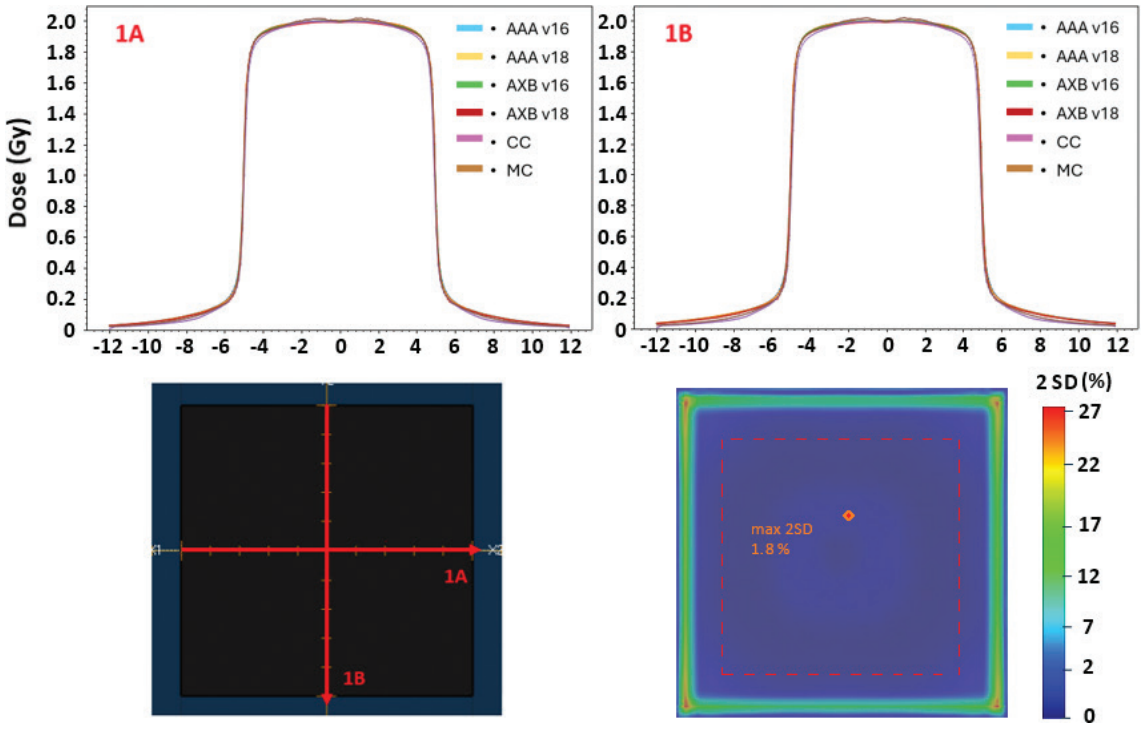


Figure 1. First row: dose profiles (1A, 1B) from a static 10 × 10 cm jaw-only field calculated using six methods. Field properties: 240 monitor units, 10 cm depth, source-to-surface distance 90 cm. Profiles 1A and 1B correspond to the x- and y-axes, respectively, according to the IEC 61217 coordinate system. Bottom row: (1) profile positions in jaw-only field (left), and (2) 2D map at 10 cm depth of the 2-standard-deviation in % of the local dose between the different dose calculation methods for the open field (right). The red point illustrates the point with the highest 2SD value within the 8 × 8 cm^2 central part of the beam, which was 1.8% and located at 0.6 and 1.7 cm off-axis in the x and y directions, respectively.

PTV were similar across all treatment sites, except head & neck, with ranges of 0.9–1.2% and 1.8–2.9% respectively. For the head & neck $2SD_{2cc}$ values, it was up to 5.1%. The $2SD_{Mean}$ and $2SD_{2cc}$ values for the region 1 cm outside the PTV were generally larger (highest values 3.1 and 18.7%, respectively) than the corresponding values for the PTVs. Overall, the results showed that

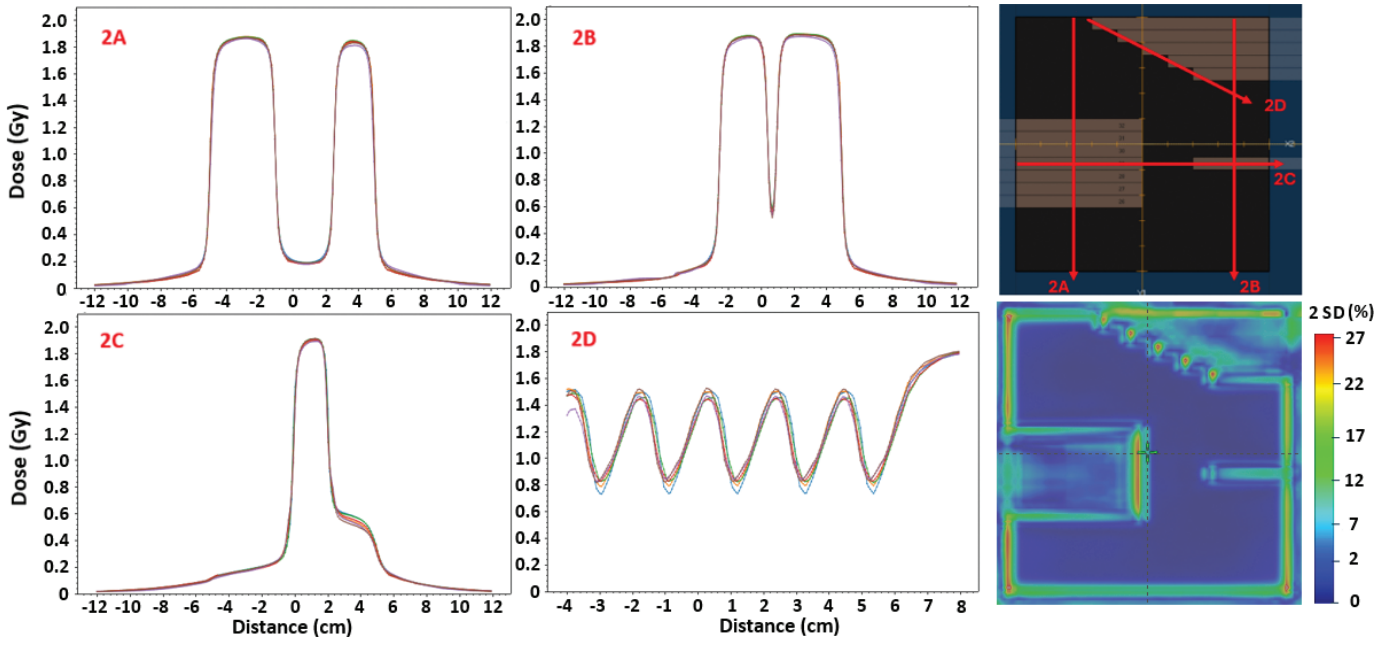


Figure 2. Dose profiles (2A-2D) from a static 10 × 10 cm jaw field with multileaf collimator (MLC), calculated using six methods. Field properties: 240 monitor units, 10 cm depth, source-to-surface distance 90 cm. Profiles 2A and 2B in the y-direction and 2C in the x-direction, according to the IEC 61217 coordinate system. Rightmost column: (1) illustration of field geometry and profile positions (top) and (2) 2D map at 10 cm depth of the 2-standard-deviation in % of the local dose between the different dose calculation methods (bottom).

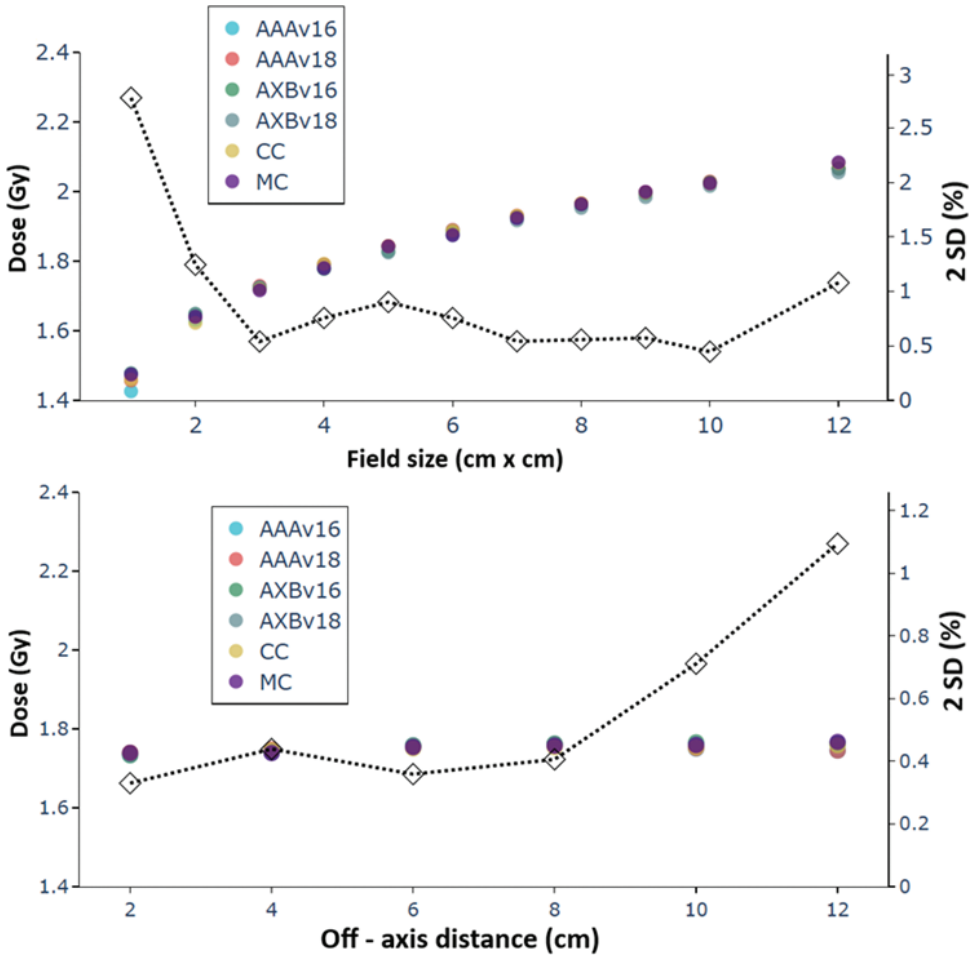


Figure 3. (Upper) Isocenter doses for static square fields of various sizes using different dose calculation methods. (Lower) Dose at 10 cm depth in the center of a 3 × 3 cm field for varying lateral shifts from isocenter toward the X2 jaw. All fields were calculated with 240 MU and 90 cm SSD. 2 standard deviations (2SD) in % of the local dose between calculation methods at each point is indicated by the diamond markers connected with a dotted line.

the influence of plan complexity on dose calculation variations was case-dependent.

Discussion and conclusion

Variations in dose distributions calculated using six different methods were investigated for static fields and VMAT plans of various complexity and treatment sites. In the static fields, higher 2SDs were observed for regions close to the MLC leaf-tips, in smaller fields and at larger off axis positions. For VMAT plans, dose variations were generally higher in regions outside the PTV compared to inside the PTV. Dose variations in the OOIs were larger than within the PTV, although the magnitude varied between patient cases and treatment sites. Variations between patient cases were generally larger than variations between complexity levels. Identifying regions where beam models and calculation algorithms diverge provides valuable insight into which anatomical regions are more prone to dose calculation uncertainties and where clinical dose estimates may be less reliable.

The results for the 10 × 10 cm^2 jaw-defined static open field showed that the 2SDs of the calculated doses were larger off axis compared the center, that is, within 0.5% in the center and 1.8% within the 8 × 8 cm^2 central part of the beam. However, the 2SDs in the static field with MLC protrusions were much larger, especially close to the MLC leaf-tip, which indicates that the off-axis variations were less dominant in the MLC field. Based on the results obtained for the static fields, it was anticipated that VMAT treatment plans with complex aperture shapes and high beam modulation would exhibit larger dose calculation variations. Hence, higher $2SD_{Mean}$ and $2SD_{2cc}$ values were expected for complex plans compared to clinical and simple plans. However, the difference in average values per treatment site was within 0.9 and 1.7% for the PTV and OOIs, respectively (Table 2). Statistically significant differences in 2SD histograms between treatment plans of different complexity levels were observed in several cases (Table 1). Differences between complexity levels were generally smaller than those between individual patient cases.

The evaluations for the open static field were consistent with the findings of Rostami et al. [7], who performed a comparative evaluation of dose calculation methods (i.e. AAA and AXB (v.16.1) in Eclipse and CC and MC in RayStation) similar to those used in the present study. Similar to their findings, for an open 10×10 cm^2 field, the dose difference between different algorithms in the isocenter was less than 1%. In addition, our study extended the evaluation to include static fields with MLC protrusions as well as VMAT treatment plans. It is worth noticing

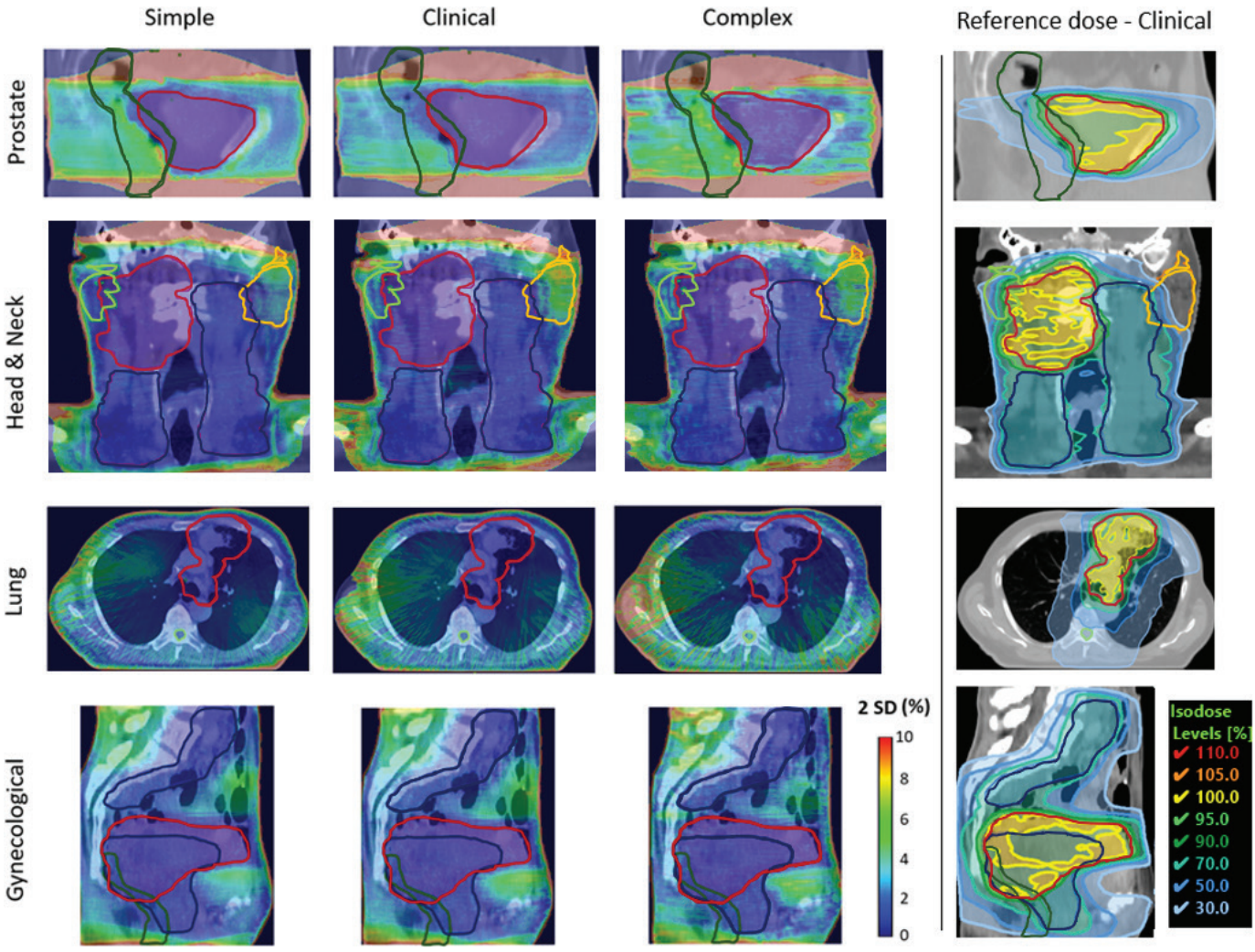


Figure 4. Representation of voxel-wise 2 standard deviations (2SD, SD with $k = 2$) distributions for one case each, prostate (plan #1), lung (plan #2), head & neck (plan #2), and gynecological (plan #3) divided between simple, clinical, and complex plans. For each voxel, the 2SD of dose for six different dose calculation algorithms normalized to the local dose is shown. Voxels with reference dose less than 5% were excluded. For each treatment site, a central slice through the planning target volume (PTV) is depicted. Coronal views are shown for head & neck cases, transversal for the lung cases, while sagittal views are presented for prostate and gynecological cases. The red contours refer to the PTVs of the primary tumors, and the blue contours indicate the PTVs of the lymph nodes. The dark green contours refer to rectum for the prostate and gynecological cases. For the lung case, the spinal cord is depicted with a light green contour. For the head & neck case, the light green and orange contours refer to the ipsilateral and contralateral parotids, respectively. The reference dose distribution for each clinical case is shown in the rightmost column.

that Rostami et al. also investigated different beam energies, flattening-filter-free beams, and varying SSDs. These factors are not included in this study but are potentially relevant, given their dependence on the beam model.

Li et al. [13] investigated how discrepancies between different dose calculation algorithms relate to treatment plan complexity for Intensity Modulated Radiotherapy (IMRT). They evaluated 53 complexity metrics and found that discrepancies among CC, MC, and AAA algorithms correlated with some of these metrics. In their analysis, metrics related to plan modulation (e.g. variation in MLC speed and acceleration) were found to have stronger correlations with differences between dose calculation algorithms than metrics related to the aperture shape, for example, the edge area metric (EAM) [17, 18]. In the present study, VMAT plans were divided into different complexity levels. It has been shown in a previous study that significant differences in EAM values exist between these complexity levels [16]. In this study, several statistically significant differences were found in the SD histograms between plans of different complexity levels.

The number of dose calculation algorithms included in this study was relatively small, which limits the strength of the standard deviation estimated in each voxel. This limitation could be addressed by effectively increasing the sample size through systematic perturbations of the existing beam configurations, for example, by introducing small variations to parameters such as the dosimetric leaf gap (DLG) or leaf transmission. As demonstrated by former studies, for example [19, 20], even an offset in the DLG as small as 1 mm can result in notable dose variations in VMAT plans.

In this study, only dose variations related to the calculation method were considered. Incorporating additional sources of uncertainty, such as those related to treatment delivery, would

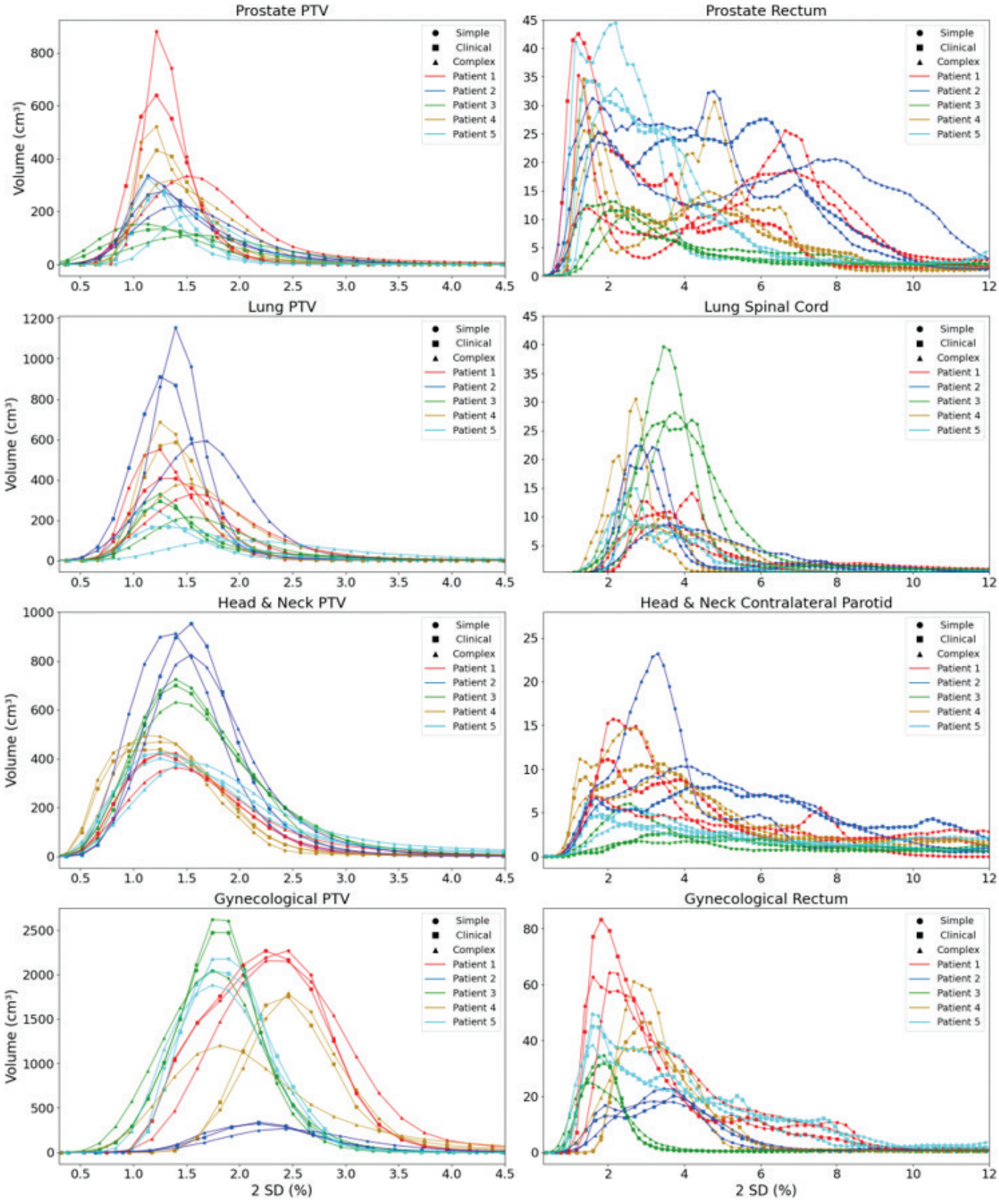


Figure 5. Two standard deviation (2SD, SD with $k = 2$) differential histograms for all cases and plans grouped by four different treatment sites. The bin widths for x-axis were 0.1% 2SD. For prostate and lung cases, the PTV encompasses the primary tumor only, whereas for head & neck and gynecological cases, lymph node PTVs are included when applicable.

Table 1. *P*-values from Wilcoxon signed-rank tests comparing the standard deviation with $k = 2$ (2SD) distribution between the clinical and simple version and between the clinical and complex version for each case. PTV and one organ of interest (OOI) were analyzed: rectum (prostate, gynecological), spinal cord (lung), contralateral parotid (head & neck).

Case	PTV: Clinical vs. simple	PTV: Clinical vs. complex	OOI: Clinical vs. simple	OOI: Clinical vs. complex
Prostate #1	< 0.05	< 0.05	0.77	0.49
Prostate #2	< 0.05	< 0.05	0.05	0.11
Prostate #3	0.14	< 0.05	0.31	0.49
Prostate #4	0.10	< 0.05	0.69	0.31
Prostate #5	0.13	< 0.05	< 0.05	0.52
Head & Neck #1	< 0.05	< 0.005	0.06	< 0.05
Head & Neck #2	< 0.05	0.92	< 0.05	0.53
Head & Neck #3	< 0.05	< 0.05	0.40	< 0.05
Head & Neck #4	0.17	< 0.05	< 0.05	< 0.05
Head & Neck #5	< 0.05	< 0.05	< 0.05	0.20
Lung #1	< 0.05	< 0.05	< 0.05	< 0.05
Lung #2	< 0.05	< 0.05	0.59	< 0.05
Lung #3	0.24	0.29	0.42	< 0.05
Lung #4	0.35	< 0.05	< 0.05	< 0.05
Lung #5	< 0.05	< 0.05	0.35	< 0.05
Gynecological #1	< 0.05	< 0.05	0.85	0.21
Gynecological #2	0.19	0.06	0.19	0.10
Gynecological #3	0.48	0.18	0.08	0.06
Gynecological #3	< 0.05	0.85	0.92	0.95
Gynecological #3	0.31	< 0.05	< 0.05	0.16

PTV: planning target volume; OOI: one organ of interest.

PTV and OOI were analyzed: rectum (prostate, gynecological), spinal cord (lung), and contralateral parotid (head & neck).

Table 2. Average values (three cases per treatment site and complexity level) for the maximum standard deviation with $k = 2$ (2SD) to a volume of 2 cm^3 ($2SD_{2cc}$) and the mean value of 2SD ($2SD_{Mean}$) inside the PTV, in the region 1 cm outside the PTV as well as for one organ (OOI) at risk per treatment site: rectum (prostate, gynecological), spinal cord (lung), and contralateral parotid (head & neck).

Prostate		Inside PTV	1 cm outside PTV	OOI
Average $2SD_{2cc}$ (Local %)	Simple	1.9	9.7	14.6
	Clinical	1.9	11.2	14.8
	Complex	2.8	13.1	15.0
Average $2SD_{Mean}$ (Local %)	Simple	0.9	2.4	4.2
	Clinical	1.0	2.3	4.4
	Complex	1.2	3.1	5.0
Head & Neck		Inside PTV	1 cm outside PTV	OOI
Average $2SD_{2cc}$ (Local %)	Simple	4.5	18.0	7.8
	Clinical	4.8	18.3	9.1
	Complex	5.1	18.7	9.3
Average $2SD_{Mean}$ (Local %)	Simple	1.1	2.0	3.3
	Clinical	1.2	2.2	4.3
	Complex	1.2	2.4	4.7
Lung		Inside PTV	1 cm outside PTV	OOI
Average $2SD_{2cc}$ (Local %)	Simple	1.8	10.2	7.2
	Clinical	2.1	13.3	7.2
	Complex	2.9	12.6	8.9
Average $2SD_{Mean}$ (Local %)	Simple	0.9	1.9	2.8
	Clinical	1.0	2.2	2.8
	Complex	1.2	2.6	3.6
Gynecological		Inside PTV	1 cm outside PTV	OOI
Average $2SD_{2cc}$ (Local %)	Simple	2.1	8.4	13.3
	Clinical	2.3	9.4	14.8
	Complex	2.5	10.5	14.8
Average $2SD_{Mean}$ (Local %)	Simple	1.1	1.8	2.9
	Clinical	1.1	1.9	3.2
	Complex	1.1	2.0	3.3

PTV: planning target volume; OOI: one organ of interest.

allow for a more complete understanding of how different factors contribute to the overall accuracy of radiotherapy dose estimation. In this context, our group has previously investigated delivery-related variations by introducing systematic and random offsets of treatment machine parameters, to simulate potential deviations during treatment delivery [16]. Results from that study showed that 2SDs within the CTV were larger for plans of higher complexity, highlighting that they are more sensitive to delivery variations. This was also the case in the current study where differences between dose calculation methods were observed between plans of different complexity for the PTV. Both studies, however, suggest that increased variations can mainly be observed in regions outside the border of the PTV, highlighting these regions as more sensitive to both calculation and delivery variations. Future analyses combining both calculation and delivery uncertainties could offer valuable insights into the cumulative effects that influence treatment quality in advanced radiotherapy.

The different calculation methods agreed well for open static fields, but larger variations were found in regions near the MLC leaf-tips, in small fields (< 2 cm) and at large off axis positions (≥ 10 cm). For VMAT plans, variations between different dose calculation methods were larger in OOIs, where 2SDs larger than 10% (normalized to local dose) were found in some cases compared the PTVs, where the 2SDs were most often within 4%. Differences between patient cases were generally larger than differences between complexity levels. These findings suggest that inter algorithm differences, thus the estimated dose calculation uncertainty, are more strongly influenced by patient-specific factors than by plan complexity.

Acknowledgments

The authors gratefully acknowledge the financial support by the King Gustaf V Jubilee Clinic Cancer Research Foundation. NACP 2026 was financially supported by the Acta Oncologica Foundation. Eric Grönlund is greatly acknowledged for sharing his expertise and providing invaluable assistance with the RayStation Treatment Planning System.

Conflict of interest

Our department has a research agreement with Siemens Healthineers.

Data availability statement

Due to patient privacy restrictions, the datasets generated and/or analyzed in this study are not publicly available. However, parts of the data could be provided by the corresponding author upon reasonable request.

Ethics declarations & trial registry information

An ethical review application was submitted for this project and assessed by the Swedish Ethical Review Authority, which determined that no formal ethical approval was required (Registry number: 2022-06173-01). The clinical data used in this study originate from previously completed treatments and pose no risk to patient health or safety. The project is classified as development and quality assurance, and no human or animal subjects were directly involved.

Author contributions

E. Terzidis, F. Nordström, M. Gustafsson, A. Karlsson, J. Götstedt, and A. Bäck conceptualized and designed the study. E. Terzidis conducted formal analyses, data curation, and visualization and drafted the original manuscript. All authors contributed to methodology development, investigation, and manuscript review and editing. F. Nordström and A. Bäck supervised the work, and A. Bäck managed the project and funding acquisition.

References

[1] Ma CMC, Chetty IJ, Deng J, Faddegon B, Jiang SB, Li J, et al. Beam modeling and beam model commissioning for Monte Carlo dose calculation-based radiation therapy treatment planning: report of AAPM Task Group 157. Med Phys. 2020;47(1):e1–18. https://doi.org/10.1002/mp.13898

[2] Stathakis S, Esquivel C, Quino LV, Myers P, Calvo O, Mavroidis P, et al. Accuracy of the small field dosimetry using the acuros XB dose calculation algorithm within and beyond heterogeneous media for 6 MV photon beams. Int J Med Phys Clin Eng Radiat Oncol. 2012;01(03):78–87. https://doi.org/10.4236/ijmpcero.2012.13011

[3] Das IJ, Ding GX, Ahnesjo A. Small fields: nonequilibrium radiation dosimetry. Med Phys. 2008;35(1):206–15. https://doi.org/10.1118/1.2815356

[4] Wolfs CJA, Swinnen ACC, Nijsten S, Verhaegen F. Should dose from small fields be limited for dose verification procedures? Uncertainty versus small field dose in VMAT treatments. Phys Med Biol. 2018;63(20):20NT01. https://doi.org/10.1088/1361-6560/aae338

[5] Liu HW, Nugent Z, Clayton R, Dunscombe P, Lau H, Khan R. Clinical impact of using the deterministic patient dose calculation algorithm Acuros XB for lung stereotactic body radiation therapy. Acta Oncol. 2014;53(3):324–9. https://doi.org/10.3109/0284186X.2013.822552

[6] Han T, Mikell JK, Salehpour M, Mourtada F. Dosimetric comparison of Acuros XB deterministic radiation transport method with Monte Carlo and model-based convolution methods in heterogeneous media. Med Phys. 2011;38(5):2651–64. https://doi.org/10.1118/1.3582690

[7] Rostami A, Neto AJC, Paloor SP, Khalid AS, Hammoud R. Comparison of four commercial dose calculation algorithms in different evaluation tests. J Xray Sci Technol. 2023;31(5):1013–33. https://doi.org/10.3233/XST-230079

[8] Reis CQM, Nicolucci P, Fortes SS, Silva LP. Effects of heterogeneities in dose distributions under nonreference conditions: Monte Carlo simulation vs dose calculation algorithms. Med Dosim. 2019;44(1):74–82. https://doi.org/10.1016/j.meddos.2018.02.009

[9] Inoue K, Matsukawa H, Kasai Y, Edamitsu K, Matsumoto K, Suetsugu Y, et al. Difference in target dose distributions between Acuros XB and collapsed cone convolution/superposition and the impact of the tumor locations in clinical cases of stereotactic ablative body radiotherapy for lung cancer. J Cancer Res Ther. 2023;19(5):1261–6. https://doi.org/10.4103/jcrt.jcrt_1740_21

[10] Bosse C, Narayanasamy G, Saenz D, Myers P, Kirby N, Rasmussen K, et al. Dose calculation comparisons between three modern treatment planning systems. J Med Phys. 2020;45(3):143–47. https://doi.

org/10.4103/jmp.JMP_111_19

[11] Fogliata A, Lobefalo F, Reggiori G, Stravato A, Tomatis S, Scorsetti M, et al. Evaluation of the dose calculation accuracy for small fields defined by jaw or MLC for AAA and Acuros XB algorithms. Med Phys. 2016;43(10):5685. https://doi.org/10.1118/1.4963219

[12] Younes T, Chatrie F, Zinutti M, Simon L, Fares G, Vieillevigne L. Optimization of the Eclipse TPS beam configuration parameters for small field dosimetry using Monte Carlo simulations and experimental measurements. Phys Med. 2023;114:103141. https://doi.org/10.1016/j.ejmp.2023.103141

[13] Li S, Luo H, Tan X, Qiu T, Yang X, Feng B, et al. The impact of plan complexity on calculation and measurement-based pre-treatment verifications for sliding-window intensity-modulated radiotherapy. Phys Imaging Radiat Oncol. 2024;31:100622. https://doi.org/10.1016/j.phro.2024.100622

[14] Yang Y, Shoniyozov K, Li T, Lockamy V, Bieda M, Teo BK, et al. Evaluation of the accuracy of an enhanced MLC leaf model for Linac-based stereotactic radiosurgery. J Appl Clin Med Phys. 2025;26(7):e70143. https://doi.org/10.1002/acm2.70143

[15] Saez J, Bar-Deroma R, Bogaert E, Cayez R, Chow T, Clark CH, et al. Universal evaluation of MLC models in treatment planning systems based on a common set of dynamic tests. Radiother Oncol. 2023;186:109775. https://doi.org/10.1016/j.radonc.2023.109775

[16] Terzidis E, Nordstrom F, Gotstedt J, Back A. Impact of delivery variations on 3D dose distributions for volumetric modulated arc therapy plans of various complexity. Med Phys. 2024;51(11):8466–81. https://doi.org/10.1002/mp.17310

[17] Gotstedt J, Karlsson A, Back A. Development and evaluation of aperture-based complexity metrics using film and EPID measurements of static MLC openings. Med Phys. 2015;42(7):3911–21. https://doi.org/10.1118/1.4921733

[18] Gotstedt J, Back A. Edge area metric complexity scoring of volumetric modulated arc therapy plans. Br J Radiol. 2021;17:124–9. https://doi.org/10.1016/j.phro.2021.02.002

[19] Lee JWL, Choi KS, Hong S, Kim YL, Chung JB, Lee D, et al. Effects of static dosimetric leaf gap on MLC-based small-beam dose distribution for intensity-modulated radiosurgery. J Appl Clin Med Phys. 2007;8(4):54–64. https://doi.org/10.1120/jacmp.v8i4.2397

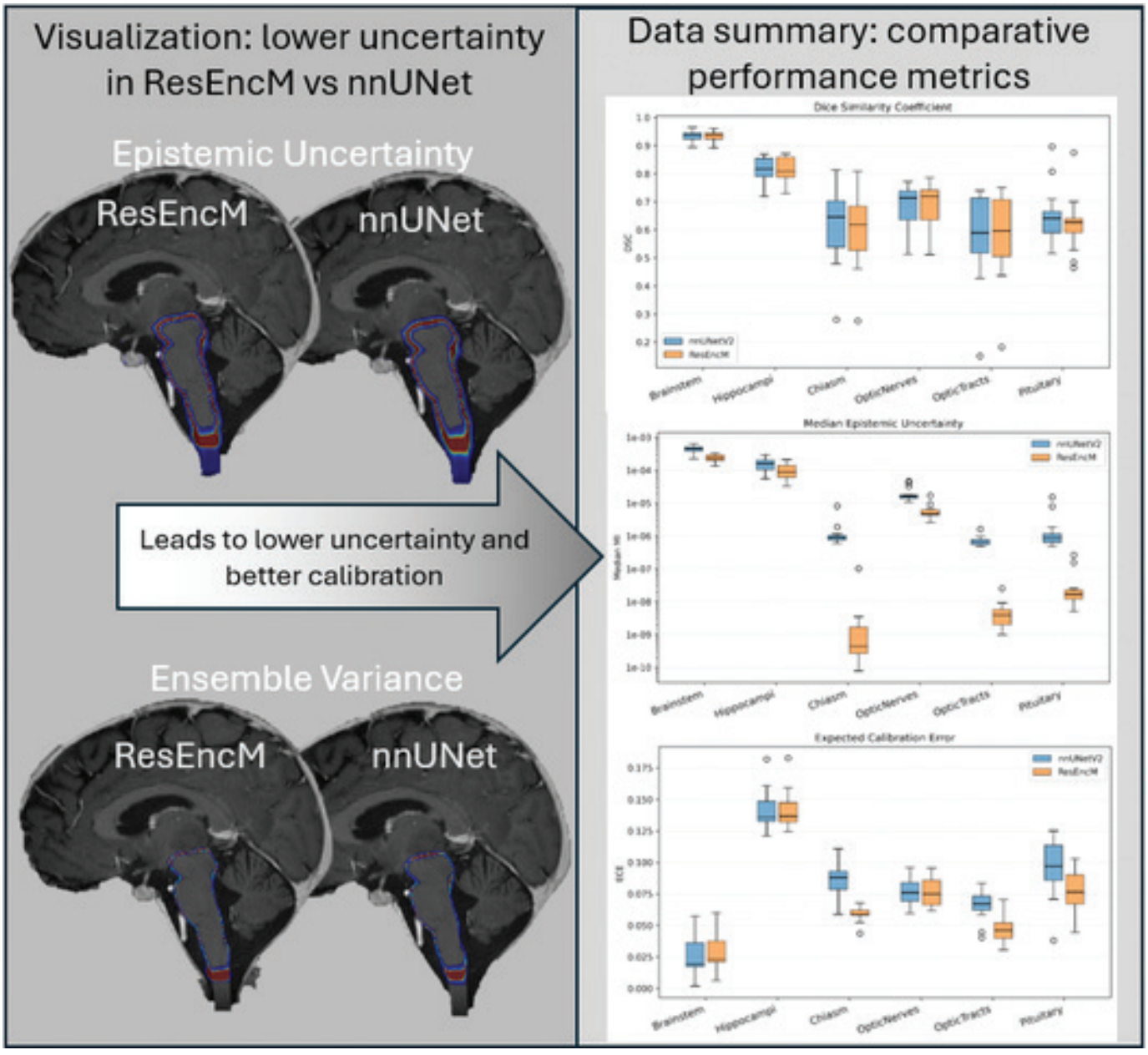
[20] Kielar KN, Mok E, Hsu A, Wang L, Luxton G. Verification of dosimetric accuracy on the TrueBeam STx: rounded leaf effect of the high definition MLC. Med Phys. 2012;39(10):6360–71. https://doi.org/10.1118/1.4752444

Improving confidence in MRI-based auto-segmentation via uncertainty assessment

Jesper Folsted Kallehauge[a,b], Jintao Ren[a,b] and Yasmin Lassen-Ramshad[a]

[a]Danish Centre for Particle Therapy, Aarhus University Hospital, Aarhus, Denmark; [b]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

ACTA ONCOLOGICA
2026, VOL. 65, 406–411
https://doi.org/10.2340/1651-226X.2026.45685



ACTA ONCOLOGICA
2026, VOL. 65, 406–411
https://doi.org/10.2340/1651-226X.2026.45685

NACP 2026

MJS PUBLISHING

SHORT REPORT

Improving confidence in MRI-based auto-segmentation via uncertainty assessment

Jesper Folsted Kallehauge[a,b], Jintao Ren[a,b] and Yasmin Lassen-Ramshad[a]

[a]Danish Centre for Particle Therapy, Aarhus University Hospital, Aarhus, Denmark; [b]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

ABSTRACT

Background and purpose: Accurate delineation of organs of interest (OOIs, also commonly referred to as organs at risk, OARs) is crucial for safe radiotherapy. While deep learning-based segmentation using convolutional neural networks has achieved high geometric accuracy, clinical translation is hindered by overconfident, uncalibrated predictions in anatomically ambiguous regions. Uncertainty quantification and model calibration are prerequisites for safe clinical workflows. This study compared the standard nnU-Netv2 against its residual-encoding variant (ResEncM), hypothesizing that ResEncM would demonstrate superior reliability and calibration while maintaining comparable geometric accuracy.

Patient/material and methods: T1-weighted contrast-enhanced MRI scans from 70 brain cancer patients were used (55 training/validation, 15 testing). Ground-truth contours for brainstem, hippocampi, chiasm, optic nerves, optic tracts, and pituitary were delineated per Danish Neuro Oncology Group guidelines. Both architectures were trained using five-fold cross-validation with identical preprocessing. Epistemic uncertainty was quantified using mutual information, and Expected Calibration Error (ECE) was computed within a 10-mm isotropic margin around reference contours.

Results: Both models achieved high geometric accuracy (brainstem dice similarity coefficient [DSC] > 0.93, hippocampi DSC > 0.81). No significant geometric differences were found for large structures. ResEncM showed significantly lower DSC for the pituitary ($p = 0.003$) and chiasm ($p = 0.018$). However, ResEncM demonstrated significantly lower epistemic uncertainty and ensemble variance across all structures ($p < 0.05$), and significantly reduced ECE for the optic chiasm, optic tracts, and pituitary.

Interpretation: Integrating a deep residual encoder into the standard U-Net framework significantly improves reliability and calibration of automated brain OOI contours while maintaining strong geometric performance. The ResEncM architecture provides a more trustworthy tool for clinical radiotherapy by reliably flagging high-uncertainty voxels, supporting confidence-aware clinical workflows.

ARTICLE HISTORY
Received 10 March 2026
Accepted 22 April 2026
Published 11 May 2026

KEYWORDS
Deep learning; image segmentation; magnetic resonance imaging; uncertainty; calibration; radiotherapy planning, computer-assisted; brain neoplasms

Introduction

Accurate delineation of organs of interest (OOIs) is crucial for safe and effective radiotherapy. In the era of artificial intelligence (AI), the field has transitioned from manual contouring – a process influenced by significant inter-observer variability [1] and inefficiency – to automated segmentation using Deep Neural Networks (DNNs) [2]. Among these, the nnU-Net framework established itself as the de facto standard, owing to its self-configuring approach that robustly handles diverse medical imaging datasets [3]. However, as the clinical integration of these models accelerates, the focus of the research community has moved from pure geometric accuracy (e.g., measured by Dice Similarity Coefficients) to model reliability, calibration, and uncertainty quantification (UQ) [4].

Despite the high geometric accuracy of standard convolutional neural networks (CNNs)-based UNets in medical image segmentation, their clinical translation is frequently hindered by a tendency to produce highly overconfident and uncalibrated predictions in anatomically ambiguous regions. To address this safety concern, recent literature heavily emphasizes UQ as a prerequisite for safe human-in-the-loop AI radiotherapy workflows [5]. UQ fundamentally distinguishes between two sources of doubt: aleatoric uncertainty, which arises from inherent data noise such as MRI artifacts or ambiguous tissue contrast, and epistemic uncertainty, which represents model ignorance and can be mitigated in part by architectural improvements or additional training data [6]. To capture this epistemic doubt, Deep Ensembles – which aggregate predictions across multiple independently trained models – have become a widely adopted strategy. They frequently demonstrate more reliable uncertainty estimation than single deterministic networks and serve as a highly robust alternative to Monte Carlo Dropout approximations [7]. Furthermore, the field is increasingly prioritizing model calibration, measured via

CONTACT Jesper Folsted Kallehauge jespkall@rm.dk Danish Centre for Particle Therapy, Palle Juul-Jensens Boulevard 99, 8200 Aarhus N, Aarhus University Hospital, Aarhus, Denmark

Supplemental data for this article can be accessed online at https://doi.org/10.2340/1651-226X.2026.45685

Expected Calibration Error (ECE) [7], to ensure that a network's output softmax probabilities accurately reflect its true likelihood of correctness. In simpler terms, calibration ensures that if the AI is 90% confident in a treatment boundary, that boundary is actually correct 90% of the time, preventing dangerous overconfidence. However, while standard U-Net architectures typically achieve high spatial overlap metrics, they often lack the internal mechanisms to maintain calibration at complex boundaries, motivating the exploration of deeper, feature-rich architectures to inherently improve model reliability [8].

Recent architectural advancements have introduced residual-encoding variants (such as ResEncM) to the standard nnU-Net framework. While residual connections are known to improve feature extraction and gradient flow, their specific impact on model certainty, predictive stability, and probability calibration in small or complex brain structures remains under-investigated.

Therefore, this study compared the standard nnUNet (v2) against its residual-encoding variant (ResEncM) [9]. We hypothesized that while ResEncM would maintain comparable geometric accuracy to the standard nnU-Netv2, it would demonstrate superior reliability and calibration.

Patients/material and methods

Patient data and image acquisition

The study included T1-weighted contrast-enhanced MRI scans from 70 adult patients with brain cancer, acquired on a 3T MRI scanner (Philips Ingenia) at a single institution. The dataset was partitioned into 55 patients for training/validation and 15 patients for independent testing. Ground-truth contours for OOIs – including the brainstem, hippocampi, chiasm, optic nerves, optic tracts, and pituitary – were delineated according to Danish Neuro Oncology Group (DNOG) consensus guidelines [10].

Model architecture and training

We configured a baseline 3D U-Net (nnU-Netv2) architecture and a residual-encoding variant (ResEncM) for this study. For both models, the input included a single modality channel (T1-weighted contrast-enhanced MRI) with an input patch size set to 128 × 128 × 128. The baseline nnU-Netv2 consisted of six depth levels (32, 64, 128, 256, 320, and 320 feature channels, respectively), utilizing two standard convolutional operations per encoder and decoder stage.

To evaluate the impact of architectural modifications, the ResEncM model variant maintained the identical input patch size, depth levels, and feature channel distribution to ensure a controlled comparison. However, its standard convolutional encoder was replaced with a Residual Encoder. Specifically, the ResEncM encoder stages utilized an increasing number of residual blocks (1, 3, 4, 6, 6, and 6 blocks across the respective depth levels) to enhance feature propagation, while the decoder was streamlined to utilize a single convolutional operation per stage.

Both architectures were trained using identical preprocessing pipelines, data augmentations, and a five-fold cross-validation strategy. All five folds contributed to the final ensemble predictions for each test patient.

Uncertainty and calibration assessment

Epistemic uncertainty was quantified voxel-wise using mutual information, and ensemble variance was derived from the variance of the softmax probabilities across the model folds. To evaluate calibration, the ECE was computed per structure using fold-averaged softmax outputs. ECE calculations were restricted to a 10-mm isotropic margin around the reference contours to focus the metric on the clinically relevant organ boundaries.

Geometric accuracy and statistical analysis

Geometric accuracy was evaluated using the Dice similarity coefficient (DSC) and 1-mm Normalized Surface Dice (1NSD). Differences in accuracy, uncertainty, and calibration metrics between nnU-Netv2 and ResEncM were assessed using paired Wilcoxon signed-rank tests, with statistical significance defined as $p < 0.05$.

Results

Geometric accuracy

Both the baseline nnU-Netv2 and the ResEncM architectures achieved high geometric accuracy across the evaluated OOIs. Median DSC exceeded 0.93 for the brainstem and 0.81 for the hippocampi for both models (Table 1). Statistical analysis revealed no significant differences in geometric accuracy (DSC and 1-mm Normalized Surface Dice) between the two architectures for the majority of the large-volume structures. However, a slight performance shift was observed for small, highly variable anatomies; the ResEncM model demonstrated significantly lower DSC and 1NSD compared to the baseline nnU-Netv2 for both the pituitary gland ($p = 0.003$) and the optic chiasm (DSC $p = 0.018$, 1NSD $p = 0.005$) (Table 1).

Uncertainty and calibration

Despite the localized reductions in spatial overlap for small structures, the integration of residual encoding yielded marked improvements in model reliability and confidence. The ResEncM architecture demonstrated consistently and significantly lower epistemic uncertainty and ensemble variance across all evaluated structures (all $p < 0.05$; see Table 1 for individual p-values) (Table 1). These reductions were most pronounced for small, low-contrast anatomies, with the largest quantitative improvements observed in the optic chiasm, optic tracts, and pituitary gland (Figure 1). Qualitative visual analysis confirmed this finding, revealing a tighter, less diffuse uncertainty profile strictly along the OOI boundaries for the ResEncM model compared to

Table 1. Comparison of geometric accuracy (DSC, 1NSD) and uncertainty/calibration metrics (Epistemic MI, Ensemble Variance, ECE) between the baseline nnU-Netv2 and the ResEncM architecture across brain organs of interest.

Metric / model	Brainstem	Hippocampi	Optic chiasm	Optic nerves	Optic tracts	Pituitary
DSC median (Range)						
nnU-Netv2	0.94 (0.89–0.97)	0.82 (0.72–0.87)	0.63 (0.28–0.81)	0.71 (0.42–0.77)	0.56 (0.15–0.74)	0.64 (0.51–0.90)
ResEncM	0.94 (0.89–0.96)	0.81 (0.73–0.87)	0.61 (0.28–0.81)	0.72 (0.45–0.79)	0.58 (0.18–0.75)	0.62 (0.47–0.88)
p-value	*0.561*	*0.561*	***0.018***	*0.389*	*0.679*	***0.003***
1NSD median (range)						
nnU-Netv2	0.88 (0.71–0.97)	0.90 (0.74–0.95)	0.77 (0.53–0.98)	0.85 (0.56–0.97)	0.86 (0.31–0.98)	0.68 (0.37–1.00)
ResEncM	0.88 (0.72–0.97)	0.92 (0.75–0.96)	0.77 (0.50–0.97)	0.86 (0.57–0.98)	0.82 (0.36–0.95)	0.63 (0.35–1.00)
p-value	*0.890*	*0.454*	***0.005***	*0.561*	*0.804*	***0.003***
Epistemic MI Median (Range)						
nnU-Netv2	4.6e-4 (2.2e-4–6.5e-4)	1.6e-4 (5.5e-5–3.0e-4)	8.8e-7 (6.0e-7–8.1e-6)	1.6e-5 (1.1e-5–4.7e-5)	6.5e-7 (4.8e-7–1.6e-6)	8.7e-7 (4.8e-7–1.5e-5)
ResEncM	2.4e-4 (1.4e-4–3.2e-4)	8.9e-5 (3.3e-5–2.1e-4)	4.4e-10 (7.7e-11–1.0e-7)	4.7e-6 (2.6e-6–1.7e-5)	3.9e-9 (9.7e-10–2.5e-8)	1.7e-8 (5.0e-9–2.6e-7)
p-value	***< 0.001***	***< 0.001***	***< 0.001***	***< 0.001***	***< 0.001***	***< 0.001***
Ens. Variance Median (Range)						
nnU-Netv2	3.2e-9 (5.8e-10–8.0e-9)	2.6e-11 (1.9e-12–1.6e-10)	4.1e-16 (1.6e-16–3.6e-14)	2.0e-13 (8.9e-14–2.1e-12)	1.6e-16 (7.6e-17–1.1e-15)	4.2e-16 (7.8e-17–3.1e-13)
ResEncM	9.8e-10 (2.7e-10–2.6e-9)	1.4e-11 (9.6e-13–7.3e-11)	2.8e-22 (8.6e-24–1.1e-17)	2.5e-14 (8.1e-15–4.4e-13)	1.7e-20 (1.4e-21–7.0e-19)	2.5e-19 (2.3e-20–7.2e-17)
p-value	***< 0.001***	***0.049***	***< 0.001***	***< 0.001***	***< 0.001***	***< 0.001***
ECE Median						
nnU-Netv2	0.02 (0.00–0.06)	0.14(0.12–0.18)	0.09(0.06–0.11)	0.08 (0.06–0.10)	0.07 (0.04–0.08)	0.10 (0.04–0.13)
ResEncM	0.02 (0.01–0.06)	0.14 (0.12–0.18)	0.06 (0.04–0.07)	0.08 (0.06–0.10)	0.05 (0.03–0.07)	0.08 (0.04–0.10)
p-value	***0.001***	*0.391*	***< 0.001***	*0.358*	***< 0.001***	***0.005***

DSC: dice similarity coefficient; 1NSD: 1-mm normalized surface dice; MI: mutual information; Ens. Variance: ensemble variance; ECE: expected calibration error.

Bold, italicized values indicate statistical significance ($p < 0.05$) based on paired Wilcoxon signed-rank tests ($n = 14$ pairs for uncertainty and ECE metrics, $n = 15$ for geometric metrics).

the standard nnU-Netv2 (Figure 2). A comprehensive comparison of epistemic uncertainty and ensemble variance maps for all six OOIs is provided in Supplementary Figures 1 and 2, respectively.

Furthermore, the ECE, computed within a 10-mm isotropic margin around the reference contours, was significantly reduced by the ResEncM architecture for the optic chiasm, optic tracts, and pituitary gland (Table 1). This indicates a much better alignment between the predicted softmax probabilities and true segmentation correctness in these challenging regions. ECE values did not show significant differences between the two models for the hippocampi and optic nerves. Conversely, the brainstem exhibited a slight but statistically significant increase in ECE with the ResEncM model ($p = 0.001$), though absolute calibration remained exceptionally high (median ECE 0.02) (Table 1). Overall, these metrics indicated that the ResEncM model provided superior calibration and drastically lower epistemic doubt for challenging structures, effectively substituting the diffuse, uncertain predictions of the baseline model with tightly calibrated contours.

Discussion and conclusion

The clinical translation of deep learning for radiotherapy segmentation requires models that are not only geometrically accurate but also robustly calibrated. In this study, we demonstrated that modifying the standard nnU-Netv2 architecture with a residual encoder (ResEncM) significantly reduced epistemic uncertainty and ECE in brain MRI segmentation without broadly compromising geometric accuracy.

Importantly, the baseline geometric accuracy achieved in this study is highly consistent with current state-of-the-art clinical benchmarks. For large, high-contrast structures like the brainstem and hippocampi, our models' median DSCs (>0.93 and >0.81, respectively) align with or exceed the performance of established commercial models and previous nnU-Net implementations, which typically report brainstem DSCs of 0.86–0.89 and hippocampal DSCs of 0.78–0.80. Similarly, for small, difficult structures like the optic chiasm (DSC ~0.61–0.63) and optic nerves (DSC ~0.71), our spatial overlap is on par with recent literature benchmarks that report DSCs ranging from 0.49 to 0.63 for the chiasm and 0.66 to 0.70 for the optic nerves [11].

Building upon this strong geometric foundation, our findings align with recent literature emphasizing the critical shift from pure geometric evaluation toward UQ to ensure clinical safety. Ren et al. [12] highlighted that standard convolutional networks often produce overconfident, uncalibrated predictions in challenging head and neck boundary regions. Similarly, our baseline nnU-Netv2 exhibited diffuse and overconfident epistemic doubt at the boundaries of small, low-contrast structures like the optic chiasm and optic tracts.

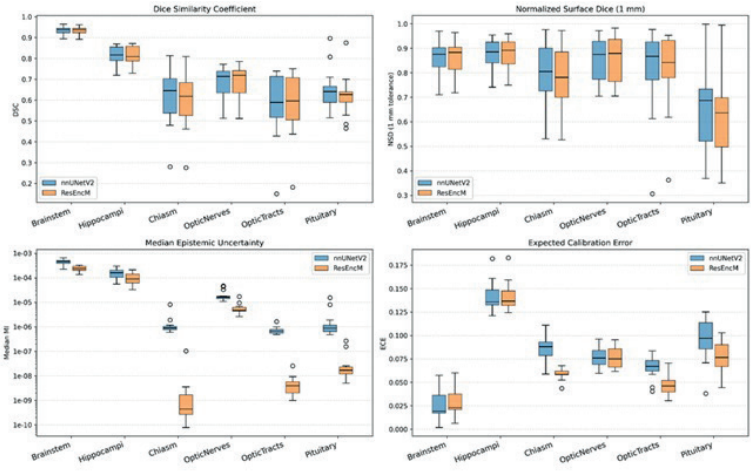


Figure 1. Quantitative impact of residual encoding on epistemic uncertainty and calibration for brain organs of interest. Box-and-whisker plots comparing prediction accuracy: Dice Similarity Coefficient (upper left) and 1 mm normalized surface Dice (upper right) and uncertainty: Epistemic Mutual Information (lower left), and Expected Calibration Error (ECE) lower right between the baseline nnU-Netv2 (blue) and ResEncM (orange) architectures for the optic chiasm, optic tracts, and pituitary gland.

By replacing the standard encoder with an asymmetric, heavy encoder, light decoder residual architecture, the ResEncM model effectively mitigated this epistemic doubt. As established by Ghaffari et al. [13] and Kumar et al. [14], utilizing an increasing number of residual blocks at the deepest spatial levels maximizes hierarchical feature extraction and prevents vanishing gradients. This architectural depth allowed the ResEncM network to build a vastly more robust internal representation of complex anatomical relationships before spatial upsampling, granting it the mathematical capacity to appropriately calibrate its confidence in ambiguous regions.

Interestingly, while ResEncM improved calibration across most structures, it yielded a significantly lower DSC for the pituitary gland and optic chiasm. This highlights a critical trade-off between geometric hedging and mathematical calibration. Standard U-Net architectures frequently achieve high spatial overlap by generating diffuse boundaries in ambiguous regions – inadvertently capturing more of the subjective ground truth while outputting untrustworthy probabilities. In contrast, the deep residual encoder forces the network to make sharp, highly calibrated decisions. While this strict decisiveness penalizes Dice scores in small, highly variable structures, it drastically reduces

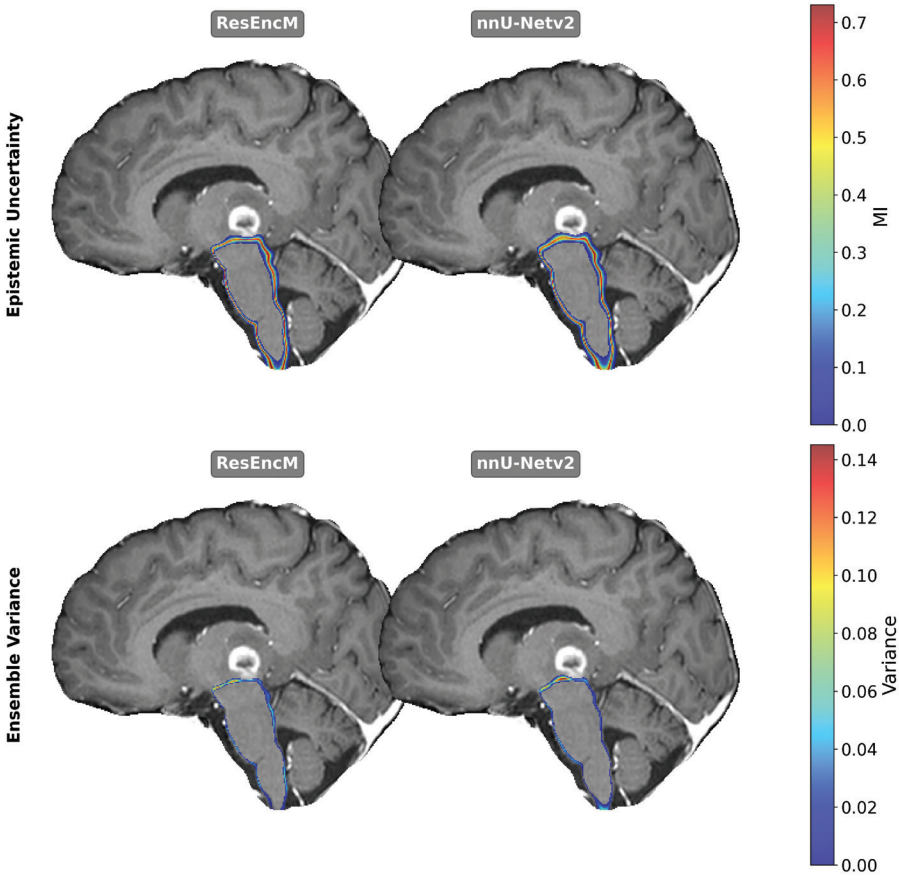


Figure 2. Comparison of ResEncM and nnUNet epistemic uncertainty for brainstem (top) and Ensemble Variance (bottom). Slightly tighter values around the brainstem were seen for ResEncM compared to nnUNet for both metrics.

epistemic uncertainty. Conversely, for structurally obvious organs like the brainstem, the baseline nnU-Netv2 was already exceptionally well-calibrated. The slight increase in ECE observed with the ResEncM model for the brainstem suggests that while deep residual encoding is vital for resolving highly ambiguous anatomies, it may lead to slight over-parameterization for simpler segmentation tasks.

A major strength of this study is the highly controlled experimental design; by utilizing the nnU-Netv2 self-configuring framework, we ensured that all preprocessing, patch sizes, and augmentations were completely identical, isolating the residual encoder as the sole driver of the improved ECE and uncertainty metrics. Furthermore, evaluating calibration within a 10-mm isotropic margin focused our analytical metrics strictly on the clinically relevant organ boundaries rather than the easily classified background.

However, several limitations regarding data quality and scope must be acknowledged. The dataset was restricted to 70 patients from a single institution, utilizing solely T1-weighted contrast-enhanced MRI scans. This single-center approach limits generalizability, as multi-institutional data with varying MRI acquisition protocols (or multi-parametric MRI) might influence baseline aleatoric noise and subsequent calibration. Additionally, the relatively small test cohort ($n = 15$) limits the statistical power of the anatomical subgroup analyses. Finally, our epistemic uncertainty estimation relies on a standard 5-fold cross-validation ensemble. While practical for standard training pipelines, an ensemble of only five models may be insufficient to fully and accurately capture the true underlying prediction variance, potentially limiting the precision of our uncertainty estimates. Notably, the current model is limited to the intracranial OOIs defined by Danish Neuro Oncology Group (DNOG) guidelines that are delineated on T1-weighted contrast-enhanced MRI; future work should incorporate additional imaging modalities to enable segmentation of the full guideline-defined structure set.

Furthermore, while this study focused on geometric and probabilistic evaluation metrics, we acknowledge that the ultimate validation of auto-segmentation tools lies in their dosimetric and clinical impact. Future studies should evaluate how uncertainty-guided workflows affect treatment plan quality, clinical acceptability of auto-contours, physician editing time, and whether uncertainty maps effectively direct clinician review to regions requiring manual correction.

While this study primarily evaluated output metrics, the disparity in calibration between the two architectures likely relates to their fundamental handling of spatial data. One limitation of the study is a missing internal feature analysis. Standard nnU-Netv2 employs sequential convolutional blocks that progressively compress information, whereas the ResEncM encoder variant utilizes residual 'identity mappings' to preserve high-resolution features from earlier layers. The superior ECE of ResEncM therefore suggests that this structural preservation of fine-grained detail may enable more reliable probability estimation at complex anatomical transition zones, where the standard architecture may be more prone to over-smoothing or information loss.

In conclusion, the integration of a deep residual encoder into the standard U-Net framework significantly improves the reliability and calibration of automated OOI contours in brain MRI. By reducing epistemic uncertainty and ECE while maintaining strong geometric performance, the ResEncM architecture provides a more transparent and trustworthy tool for clinical radiotherapy. By reliably flagging high-uncertainty voxels, this approach supports a safe, 'confidence-aware' workflow, effectively directing the clinician's manual review to areas where the AI's internal doubt is justified.

Acknowledgements

This work was supported by the DCCC Brain Tumor Center (via a grant from The Danish Cancer Society No. R295-A16770). This study was also supported by The Novo Nordisk Foundation (grant number NNF195A0059372). NACP 2026 was financially supported by the Acta Oncologica Foundation.

Conflicts of interest

The authors report there are no competing interests to declare.

Data availability statement

The clinical datasets analyzed during the current study are not publicly available due to patient privacy and institutional data protection policies. However, the trained segmentation models and the evaluation scripts used in this study are available from the corresponding author upon reasonable request.

Ethics declarations & trial registry information

This study was conducted as a local quality assurance project aimed at evaluating and improving the accuracy of auto-segmentation workflows at the Danish Centre for Particle Therapy. In accordance with Danish legislation, retrospective quality assurance studies using existing clinical data do not require formal ethics committee approval. All patient data were handled in accordance with institutional data protection policies.

Author contributions

JFK was responsible for data collection, model training, and data analysis, and drafted the initial manuscript. JR provided scientific oversight and performed a critical review of the manuscript. YL performed the label segmentation and was responsible for patient clinical management. All authors reviewed, edited, and approved the final manuscript.

References

[1] Lorenzen EL, Kallehauge JF, Byskov CS, Dahlrot RH, Haslund CA, Guldberg TL, et al. A national study on the inter-observer variability in the delineation of organs at risk in the brain. Acta Oncol. 2021;60(11):1548–54. https://doi.org/10.1080/0284186X.2021.1975813

[2] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. CoRR. 2015;abs/1505.04597:9351. https://doi.org/10.1007/978-3-319-24574-4_28

[3] Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203–11. https://doi.org/10.1038/s41592-020-01008-z

[4] Wahid KA, Ahmed S, Glerean E, Heukelom J, van der Heide UA, Fuller CD, et al. Artificial intelligence uncertainty quantification in radiotherapy applications – a scoping review. Radiother Oncol. 2024;201:110542. https://doi.org/10.1016/j.radonc.2024.110542

[5] Korreman SS, Ren J. Understanding and leveraging uncertainties in autosegmentation for radiotherapy. BJR AI. 2025;2(1):ubaf013. https://doi.org/10.1093/bjrai/ubaf013

[6] Hüllermeier E, Waegeman W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Mach Learn. 2021;110:457–506. https://doi.org/10.1007/s10994-021-05946-3

[7] Ovadia Y, Fertig E, Ren J, Nado Z, Sculley D, Nowozin S, et al. Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. arXiv. 2019. [cited 2026 May 1]. Available from: https://arxiv.org/abs/1906.02530

[8] Mehrtash A, Wells WM, Tempany CM, Abolmaesumi P, Kapur T. Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. IEEE Trans Med Imaging. 2020;39(12):3868–78. https://doi.org/10.1109/TMI.2020.3006437

[9] Isensee F, Wald T, Ulrich C, Baumgartner M, Roy S, Maier-Hein K, et al. nnU-Net revisited: a call for rigorous validation in 3D medical image segmentation. arXiv. 2024. [cited 2026 May 1]. Available from: https://arxiv.org/abs/2404.09556

[10] Danish Neuro Oncology Group (DNOG) guidelines. [cited 2026 May 1]. Available from: https://dnog.dk/onewebmedia/files/Retningslinier%20PDF/DNOG%202023%20Retningslinjer%20for%20straalebehandling.pdf

[11] Turcas A, Leucuta D, Balan C, Clementel E, Gheara C, Kacso A, et al. Deep-learning magnetic resonance imaging-based automatic segmentation for organs-at-risk in the brain: accuracy and impact on dose distribution. Phys Imaging Radiat Oncol. 2023;27:100454. https://doi.org/10.1016/j.phro.2023.100454

[12] Ren J, Teuwen J, Nijkamp J, Rasmussen M, Gouw Z, Eriksen JG, et al. Enhancing the reliability of deep learning-based head and neck tumour segmentation using uncertainty estimation with multimodal images. Phys Med Biol. 2024;69(16):165018. https://doi.org/10.1088/1361-6560/ad682d

[13] Ghaffari M, Sowmya A, Oliver R. Automated brain tumour segmentation using cascaded 3D densely-connected U-Net. In: Crimi A, Bakas S, editors. Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries. BrainLes 2020. Cham: Springer; 2021. p. 467–77.

[14] Kumar A, Ghosal P, Kundu SS, Mukherjee A, Nandi D. A lightweight asymmetric U-Net framework for acute ischemic stroke lesion segmentation in CT and CTP images. Comput Methods Programs Biomed. 2022;226:107157. https://doi.org/10.1016/j.cmpb.2022.107157

ACTA ONCOLOGICA
2026, VOL. 65, 557–560
https://doi.org/10.2340/1651-226X.2026.45692

NACP 2026

MJS PUBLISHING

LETTER

Starshot field configuration in the linac isocentre size determination: a comparison of film and electronic portal imaging device measurements

Ari-Pekka Honkanen[a] and Antti Kulmala[b]

[a]Comprehensive Cancer Center, Helsinki University Hospital, Helsinki, Finland; [b]Clinical Research Institute HUCH Ltd., Helsinki, Finland

ARTICLE HISTORY
Received 10 March 2026
Accepted 06 June 2026
Published 24 June 2026

KEYWORDS
External radiotherapy; quality assurance; winston-lutz; starshot; isocentre

Introduction

The mechanical precision of the treatment unit is paramount in the successful administration of radiotherapy. This is particularly true for stereotactic radiosurgery where sub-millimetre geometric accuracy is desirable [1, 2]. Periodic quality assurance (QA) is a must to ensure that the unit conforms to the requirements. Vendor-provided machine performance tests can ease the burden of periodic QA, but they do not remove the need for independent verification [3–6].

Two classical methods to assess the mechanical stability of the radiation isocentres of the external photon therapy unit are the starshot and Winston-Lutz tests. In the starshot test, a radiographic film is irradiated with narrow beams at multiple gantry/collimator/couch angles. The stability of the rotation axis is assessed by examining the intersection of the beams [7]. The starshot method can be automated [8], is magnetic resonance imaging (MRI) -compatible and applicable to treatment units without electronic portal imaging devices (EPID). The Winston-Lutz test involves taking portal images of a metal bead positioned at the estimated isocentre and evaluating its imaged position with respect to the radiation field edges as a function of rotation angle [9]. Traditionally the Winston-Lutz test was irradiated onto film but nowadays EPID may be used for a fully digital workflow and to extract additional information from it [10–12].

In this work, we compare an EPID-based Winston-Lutz test using a starshot irradiation pattern with the film-based starshot test in the determination of isocentre size and shape in the transaxial plane. We examine several different treatment units, collimations and beam types.

Material and methods

Three different treatment units were investigated: a Varian TrueBeam (with a Millenium multileaf collimator [MLC]), an Elekta Versa HD (Agility MLC) and a Varian Halcyon (dual-layer MLC). With the TrueBeam we measured its flattened and flattening filter free 6 MV photon beams (6X-FLT and 6X-FFF, respectively) with jaw and MLC collimations. This specimen was particularly interesting because the focal point of its 6X-FFF beam had a small lateral/transversal misalignment, which results in an approx. 1.0 mm central axis (CAX) deviation of the jaw-collimated crossplane beam profile at the isocentre.

With the Versa HD we measured flattened 6 and 10 MV beams (6X-FLT and 10X-FLT, respectively) collimated with a combination of the MLC and jaws. With the Halcyon the dual-layer MLC collimated flattening filter free 6 MV beam (6X-FFF) was measured. The collimations for these units were dictated by their technical properties as Versa HD has only one set of jaws and Halcyon is jawless.

The film measurements were done using an approx. A4-sized piece of GafChromic EBT4 film sandwiched between two slabs of solid water. The phantom was positioned at the isocentre using the lasers so that the film was in the transaxial plane. The film was irradiated at 15 gantry angles in 24° intervals starting from 186° (in IEC61217 scale). The transaxial width of the fields was nominally 1 cm at the isocentre. Each angle was irradiated twice at the collimator rotations of 90° and 270° with 150 MU each to counteract possible asymmetries in collimation. The film was digitised twice to minimise the errors in the digitisation process. A self-made FORTRAN programme was used to determine the field CAXs on the film. The isocentre was taken to be the point in the middle of the CAX intersection from which the deviations of the individual CAXs were calculated.

The EPID measurements were performed using a Winston-Lutz phantom provided in the Elekta Synergy Basic Calibration Kit MRT 15991. The phantom consisted of a metal bead approx. 8 mm in diameter embedded into a cylindrical acrylic rod. The bead was positioned to the isocentre using kilovolt cone beam computed tomography (CBCT)-imaging. Portal images of the

CONTACT Ari-Pekka Honkanen ari-pekka.honkanen@hus.fi Comprehensive Cancer Center, Helsinki University Hospital, P.O. Box 180, 00029 Helsinki, Finland

Supplemental data for this article can be accessed online at https://doi.org/10.2340/1651-226X.2026.45692

bead were acquired with the EPID using 10 MU square fields at 15 gantry angles in 24° intervals starting from 186°. The nominal field size was 2 × 2 cm at the isocentre. At each gantry angle the images were taken at five collimator rotations divided evenly over the rotational range (TrueBeam and Versa HD: 0°, 72°, 144°, 216°, 288°; Halcyon: 0°, 45°, 90°, 270°, 315°). The portal images were fed into a self-made Python programme, which automatically determines the centre positions of the fields and the bead. Using the bead as a fixed reference point, the CAXs were determined for each gantry angle from the mean of the field centres over the collimator rotations. The isocentre was defined as the point with the smallest sum of the squared distance from it to all the CAXs projected into 3D space. Only the transversal components of the CAXs and isocentre were considered for the further analysis as the longitudinal one has no counterpart in the film measurements. The technical details of the computer programmes are presented in the Supplementary material.

Results

The measured transversal deviations of CAX from their estimated isocentres are presented in Figure 1. Except for the jaw-collimated 6X-FFF of the TrueBeam, the two methods yield similar deviations for all the tested collimations, beam energies and units. These are quantified in Table 1 where means, standard deviations, and maxima of the differences between the methods are presented. Besides the exception, the mean of differences over all the gantry angles were within 0.08 mm for all the beams and collimations. This is of the same order as their calculated standard errors. The differences between the methods were within 0.20 mm for any particular gantry angle.

The exception to the trend, the jaw-collimated 6X-FFF of the TrueBeam, exhibits systematically larger CAX deviations (0.2 mm on average) in the film measurement than in the EPID one. This is reflected in the larger maximum difference of 0.37 mm, but the standard deviation of 0.07 mm is in line with the other measurements. The MLC-collimated 6X-FFF exhibits no such systematic difference between the methods. However, compared to the 6X-FLT beams of the TrueBeam, the CAX deviations are clearly larger for 6X-FFF: the mean deviations were 1.1–1.3 mm for the jaw-collimated 6X-FFF and 0.6 mm for the MLC-collimated 6X-FFF in comparison to 0.2–0.3 mm and 0.1–0.2 mm deviations of the jaw- and MLC-collimated 6X-FLT beams, respectively.

Discussion and conclusion

The measured transversal CAX deviations from the estimated isocentre between the film-based starshot and EPID-based Winston-Lutz methods were found to agree within 0.2 mm for every gantry angle for all but one beam. This is in accordance with the previously reported accuracy of the Winston-Lutz test with a similar number of sampled gantry angles [10].

The observed 0.2 mm systematic difference between the methods in the jaw-collimated 6X-FFF of the TrueBeam is most likely owing to the differences in the collimation and the resulting geometric magnification of the focal spot misalignment. In the film test the beams were collimated with the proximal jaws only whereas in the EPID measurements both proximal and distal jaws were used. Using basic geometry, a focal spot shift is magnified at the isocentre by a factor of $\frac{SAD}{d} - 1$, where SAD is

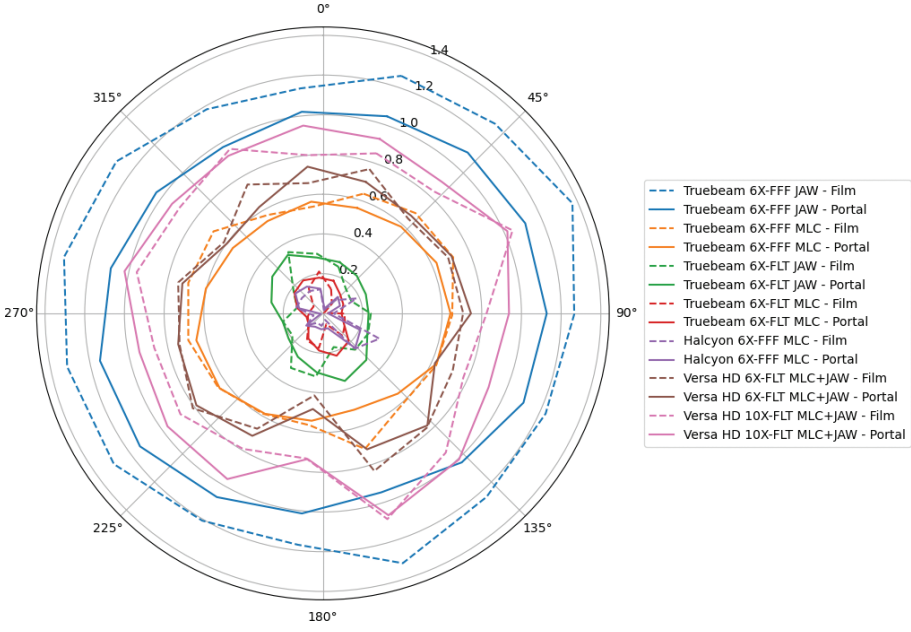


Figure 1. The deviation of the collimator rotation axis/beam central axis from the determined isocentre as a function of the gantry angle. The polar coordinate presented at the outer rim of the plot represents the gantry angle in IEC61217 scale in degrees and the radial coordinate the distance of the axis from the isocentre in mm. The solid lines are the results of the EPID measurements and the dashed lines the film measurements. The different tested beams are identified with colors and they are indicated in the legend on the right. EPID: electronic portal imaging devices.

Table 1. The means, standard deviations and maxima of differences between centre axis deviations from the isocentre between the film and electronic portal imaging measurements over the 15 sampled gantry angles.

Unit, beam, collimation	Mean difference (mm)	Standard deviation (mm)	Maximum difference (mm)
Halcyon 6X-FFF MLC	–0.006 ± 0.011	0.04	0.11
Truebeam 6X-FFF JAW	–0.20 ± 0.02	0.07	0.37
Truebeam 6X-FFF MLC	–0.06 ± 0.02	0.06	0.20
Truebeam 6X-FLT JAW	0.04 ± 0.02	0.07	0.18
Truebeam 6X-FLT MLC	0.041 ± 0.015	0.06	0.17
Versa HD 6X-FLT MLC+JAW	–0.01 ± 0.02	0.06	0.13
Versa HD 10X-FLT MLC+JAW	0.07 ± 0.02	0.06	0.17

the source-axis distance and *d* is the distance of the midpoint of the collimation device from the source. For TrueBeam SAD = 100 cm and midpoint distances of the proximal jaws, distal jaws and the MLC are 31.9, 40.6, and 50.9 cm, respectively. Therefore, the geometric magnification factors for the proximal jaws, (mean of) proximal+distal jaws, and MLC are 2.13, 1.80, 0.96, respectively. Assuming a focal spot shift of 0.6 mm, this would lead to the respective CAX deviations of 1.3, 1.1, and 0.6 mm projected at the isocentre, which are in accordance with the measured values.

The results indicate that both tests can accurately determine the CAX deviations of a treatment unit. However, the choice of collimation can affect to the results of periodic QA and possibly to the frequency of the service actions triggered by them. If the MLC is the primary mode of collimation, then performing QA using jaw-collimated beams can lead to an overly pessimistic estimation of the clinical performance of the unit and cause unnecessary maintenance and downtime. On the other hand, the observed differences between the MLC- and jaw-collimated beams can shed light on the nature of the CAX deviations and thus aid in their troubleshooting.

In conclusion, we have compared an EPID-based Winston-Lutz test using a starshot irradiation pattern to a film-based starshot test in the determination of the isocentre size and shape of external radiotherapy units. In terms of the measured transversal CAX deviations from the estimated isocentre, the methods were found to agree within 0.2 mm in the typical case. However, systematic differences were observed in the presence of the focal spot misalignment. These are most likely attributed to differing geometric magnifications at the isocentre which depend on the used collimation device. This may impact the interpretation of QA results and maintenance actions taken based on them.

Acknowledgements

This work was financially supported by the Comprehensive Cancer Center of Helsinki University Hospital. NACP 2026 was financially supported by the Acta Oncologica Foundation.

Disclosure statements

The authors report that there are no competing interests to declare.

Data availability statement

The data are available from the authors upon a reasonable request.

Ethics declarations & trial registry information

Not applicable to this work.

Authors' contributions

Both authors designed the study. The film measurements, digitalisation, processing and development of the film analysis software was performed by AK. The digital portal image measurements, processing and the development of the portal image analysis software was done by APH. The data analysis and drafting of the manuscript was carried out by APH. Both authors participated in the finalisation of the manuscript.

References

[1] Schmitt D, Blanck O, Gauer T, Fix MK, Brunner TB, Fleckenstein J, et al. Technological quality requirements for stereotactic radiotherapy. Strahlenther Onkol. 2020;196:421–33. https://doi.org/10.1007/s00066-020-01583-2

[2] Klein E, Hanley J, Bayouth J, Yin F-F, Simon W, Dresser S, et al. Task group 142 report: quality assurance of medical accelerators. Med Phys. 2009;36:4197–212. https://doi.org/10.1118/1.3190392

[3] Clivio A, Vanetti E, Rose S, Nicolini G, Belosi MF, Cozzi L, et al. Evaluation of the machine performance check application for TrueBeam Linac. Rad Oncol. 2015;10:97. https://doi.org/10.1186/s13014-015-0381-0

[4] Kron T, Lehmann J, Greer PB. Dosimetry of ionising radiation in modern radiation oncology. Phys Med Biol. 2016;61:R167. https://doi.org/10.1088/0031-9155/61/14/R167

[5] Barnes MP, Greer PB. Evaluation of the truebeam machine performance check (MPC) geometric checks for daily IGRT geometric accuracy quality assurance. J Appl Clin Med Phys. 2017;18:200–6. https://doi.org/10.1002/acm2.12064

[6] Li Y, Netherton T, Nitsch PL, Gao S, Klopp AH, Balter PA, et al. Independent validation of machine performance check for the Halcyon and TrueBeam linacs for daily quality assurance. J Appl Clin Med Phys. 2018;19:375–82. https://doi.org/10.1002/acm2.12391

[7] Wootton P, Almond PR, Holt JG, Hughes DB, Jones D, Karzmark CJ, et al. Code of practice for xray therapy linear accelerators. Med Phys. 1975;2:110. https://doi.org/10.1118/1.594133

[8] González A, Castro I, Martínez JA. A procedure to determine the radiation isocenter size in a linear accelerator. Med Phys. 2004;31:1489–93. https://doi.org/10.1118/1.1755491

[9] Lutz W, Winston KR, Maleki N. A system for stereotactic radiosurgery

with a linear accelerator. Int J Radiat Oncol Bio Phys. 1988;14: 373–381. https://doi.org/10.1016/0360-3016(88)90446-4

[10] Du W, Johnson JL, Jiang W, Kudchadker RJ. On the selection of gantry and collimator angles for isocenter localization using Winston-Lutz tests. J Appl Clin Med Phys. 2016;17:167–78. https://doi.org/10.1120/jacmp.v17i1.5792

[11] Calvo-Ortega J-F, Moragues-Femenía S, Laosa-Bello C, José-Maderuelo SS, Casals-Farran J. A closer look at the conventional Winston-Lutz test: analysis in terms of dose. Rep Pract Onc Rad. 2019;24:421–7. https://doi.org/10.1016/j.rpor.2019.07.003

[12] Chao H-Y, Li W-T, Chang C-C, Tsang Y-W. Evaluating the linear accelerator accuracy: integrating Winston-Lutz measurement method and mathematical calculations. Rad Phys Chem. 2024;217:111504. https://doi.org/10.1016/j.radphyschem.2023.111504

ACTA ONCOLOGICA
2026, VOL. 65, 609–613
https://doi.org/10.2340/1651-226X.2026.45700

NACP 2026

MJS PUBLISHING

LETTER TO THE EDITOR

Preliminary feasibility of pre-treatment F-18-PSMA-1007 PET/CT in dose prediction for Lu-177-PSMA-I&T therapy

Suvi Kokkonen[a,b], Vappu Reijonen[a], Eero Hippeläinen[b,c], Veera Ahtiainen[a], Jukka Schildt[c], Outi Sipilä[c], Sauli Savolainen[b] and Mikko Tenhunen[a]

[a]Cancer Center, Helsinki University Hospital, Helsinki, Finland; [b]Department of Physics, University of Helsinki, Helsinki, Finland; [c]Clinical Physiology and Nuclear Medicine, Helsinki University Hospital, Helsinki, Finland

Introduction

In metastatic castration-resistant prostate cancer (mCRPC), the treatment options are limited to palliative approaches, including targeted radionuclide therapy (TRT) using Lutetium-177 prostate-specific membrane antigen ([177Lu]Lu-PSMA) [1, 2]. Eligibility for [^{177}Lu]Lu-PSMA therapy is commonly assessed with a pre-treatment Fluorine-18 [^{18}F]F-PSMA positron emission tomography/computed tomography (PET/CT) to confirm sufficient PSMA accumulation in tumor lesions [3].

TRT is typically administered with fixed activity, not accounting for interpatient variability in biodistribution and biokinetics, leading to potential under- or overtreatment in some patients [4, 5]. The kidneys are considered critical organs in [^{177}Lu]Lu-PSMA therapies, in addition to red bone marrow, salivary, and lacrimal glands. Patient-specific dosimetry is recommended [6].

There is significant variation in the reported dose estimates, which may be not only related to patient characteristics but also to the influence of heterogenous methodological factors [7]. Post-treatment dosimetry in [^{177}Lu]Lu-PSMA therapy is typically performed using single-photon emission computed tomography/computed tomography (SPECT/CT), but the imaging time points, acquisition and reconstruction protocols, and segmentation methods vary between centers. There are initiatives for harmonization [8–10]. Several studies have addressed the predictive power of pre-therapy PSMA PET imaging with promising, yet varying results [11–13]. No established method exists to use pre-therapy PET imaging for individualized dose planning in PSMA TRT.

Recent studies report correlations between the [^{177}Lu]Lu-PSMA absorbed dose to tumor, the pre-treatment PSMA PET/CT uptakes, and treatment response, suggesting the treatment response could be refined with a dosimetric approach [14, 15]. Preliminarily, we hypothesize that pre-treatment PSMA PET/CT may provide quantitative information to predict absorbed doses. Exploration of the technical factors is a mandatory step preceding this, and a pilot approach was taken.

ARTICLE HISTORY
Received 10 March 2026
Accepted 28 June 2026
Published 22 July 2026

KEYWORDS
Castration-resistant prostate cancer; PSMA-PET; [177Lu]Lu-PSMA radioligand therapy; theranostic

Patients/material and methods

Patient data

This retrospective feasibility study included five mCRPC patients who underwent both [^{18}F]F-PSMA ([^{18}F]F-PSMA-1007) pre-treatment PET/CT imaging and [^{177}Lu]Lu-PSMA ([^{177}Lu]Lu-PSMA-I&T) therapy followed by two time-point post-treatment SPECT/CT imaging at the Hospital District of Helsinki (HUS).

[^{177}Lu]Lu-PSMA therapy was carried out with an administered activity of 7.4 GBq (±5%) per cycle. Dosimetry results from the first therapy cycle are studied.

In total, 10 kidney and six bone lesion volumes of interest (VOI) were analyzed. The patient-specific administered activities and CT-based volumes of kidneys and bone lesions are summarized in Table 1. The methodological workflow is presented in Supplementary Appendix A.

Imaging and reconstruction protocols

PET/CT imaging was performed approximately 90 min post [^{18}F]F-PSMA injection. Two SPECT/CT scans were taken approximately 24 h and a week post-treatment [6].

For PET/CT, either GE Healthcare Discovery MI (GE Healthcare, Chicago, IL, USA) or Siemens Biograph mCT 64 R4 (Siemens Healthcare, Erlangen, Germany) was used with EARL2-compliant reconstructions [16]. For SPECT/CT, Siemens SYMBIA pro.Specta Q3 (Siemens Healthcare, Erlangen, Germany) was used with Ordered Subsets Conjugate Gradient Minimizer (OSCGM)

CONTACT Suvi Kokkonen suvi.e.kokkonen@hus.fi HUS Cancer Center, Paciuksenkatu 3, 00290 Helsinki, Finland

Supplemental data for this article can be accessed online at https://doi.org/10.2340/1651-226X.2026.45700

Table 1. Administered activities of [^{18}F]F-PSMA for pre-treatment PET/CT and [^{177}Lu]Lu-PSMA treatment, including the CT-based segmented volumes for kidneys, denoted as K_L for left and K_R for right, and bone lesions, denoted as B_L.

	[^{18}F]F-PSMA	PET/CT V_{CT} [mL]			[^{177}Lu]Lu-PSMA	SPECT/CT V_{CT} [mL]		
Patient #	A [MBq]	K_L	K_R	BL	A [GBq]	K_L	K_R	BL
1	219	147	117	3.8	7.5	163	160	3.7
2	301	249	236	11.2	7.6	258	240	10.4
3 (1)	219	141	131	16.4	7.3	150	164	6.3
3 (2)				9.6				6.6
4	213	121	145	18.0	7.1	132	145	20.8
5	244	116	147	0.5	7.4	119	168	0.9
Median (range)	219 (213–301)	143 (116–249)		10.4 (0.5–18.0)	7.4 (7.1–7.6)	165.5 (119–258)		6.5 (0.9–20.8)

PSMA: Prostate-specific membrane antigen; PET/CT: Positron emission tomography/computed tomography; V_{CT}: CT-based delineated volume; SPECT: Single-photon emission computed tomography; K_L: Left kidney; K_R: Right kidney; B_L: Bone lesion.

reconstruction [17]. Imaging and reconstruction parameters are listed in Appendix B.

Segmentation

The kidneys and bone lesions were segmented separately on the PET/CT and SPECT/CT images, using CT- and gradient-based approaches. CT-based kidney segmentation was performed using MIM Software's ProtegeAI+ workflow, while lesions were manually segmented on CT [18]. Gradient-based segmentation for both kidneys and lesions was performed using the PET/SPECT intensity gradient-based PET Edge tool in MIM Software [19]. Only lesions clearly visible in PET, SPECT, and CT images were included.

Recovery coefficients

Recovery coefficient (RC) correction was applied to activity uptakes on CT-based bone lesion volumes to improve post-therapy SPECT scan's quantitative accuracy. RCs for SPECT's OSCGM reconstruction were determined using the standard International Electrotechnical Commission (IEC) Body Phantom from the National Electrical Manufacturers Association (NEMA) [20]. The phantom spherical inserts were filled with ^{177}Lu water solution (6.6 ± 0.2 MBq/g), followed by SPECT/CT scan using Siemens Pro.Specta Q3 [21]. Detailed imaging parameters are found in Supplementary Appendix B.

Dosimetry and dose prediction

Post-therapy absorbed doses were determined based on two-time-point SPECT/CT imaging and implemented in MIM Software [5].

Pre-treatment dose predictions were derived from PET/CT using an experimental dosimetry module (Madsen Dosimetry Estimator, MIM Software) [22], which extrapolates absorbed dose from pre-therapy activity distributions under specified effective half-life assumptions, detailed below.

Three different effective half-lives for kidneys and bone lesions were tested: patient-specific values observed from post-therapy SPECT/CT dosimetry (CT-based volumes and OSCGM reconstruction), and two literature-based values by Schuchardt et al. (33 h for kidneys and 43 h for bone lesions) [23], and Karimzadeh et al. (39 h for kidneys and 56 h for bone lesions) [24]. The patient-specific post-therapy dosimetry results and effective half-lives were considered as the reference, while literature half-life-based predictions were evaluated against these to assess the feasibility of dose prediction without the prior knowledge of individual kinetics.

Results

Recovery coefficients

SPECT recovery coefficients were modeled using monoexponential fits. For OSCGM reconstruction, the recovery curve was:

$$RC = 1.090 - 1.607e^{-0.052D}$$

where D is the sphere diameter (mm). The RC measurements and the model fit are presented in Supplementary Appendix C.

Post-therapy dosimetry using SPECT/CT

The kidneys' median absorbed dose using OSCGM reconstruction and CT-based segmentation was 3.1 Gy (range 2.0–5.9 Gy). For bone lesions, the median absorbed dose was 33.8 Gy (range 10.4–68.9 Gy). Patient-specific results can be found in Table 2 (under D_{ref}).

Using OSCGM reconstruction and gradient-based segmentation, the kidney median absorbed dose was 2.7 Gy (range 1.8–8.9 Gy), and the median bone lesion absorbed dose was 20.6 Gy (6.5–44.5 Gy).

Pre-therapy dose estimation using PET/CT

Table 2 summarizes the key findings: PET/CT-based absorbed dose predictions for kidneys and bone lesions using literature effective half-lives [23], with SPECT/CT-based observed absorbed doses and corresponding effective half-lives, using current local clinical methodology. Here, the literature half-lives [23] were chosen as the study included more patients (51 patients in [23] vs. 16 in [24]). The median ratio between

Table 2. Absorbed dose predictions ($D_{Schuchardt}$) derived from pre-treatment PET/CT scan, using CT-based segmentation and literature effective half-lives [23] for kidneys ($t_{Schuchardt}$ = 33 h) and bone lesions ($t_{Schuchardt}$ = 43 h).

Kidneys								
	PET/CT		SPECT/CT				Ratio of doses from PET/SPECT	
	$D_{Schuchardt}$ [Gy]		D_{ref} [Gy]		t_{ref} [h]		$D_{Schuchardt}/D_{ref}$	
Patient	L	R	L	R	L	R	L	R
1	7.5	6.6	3.1	3.1	30	32	2.4	2.1
2	5.0	5.8	2.0	2.4	28	29	2.5	2.4
3	4.5	2.8	3.7	4.3	35	30	1.2	0.7
4	5.5	5.5	2.0	2.1	32	30	2.8	2.6
5	4.6	6.0	4.1	5.9	29	29	1.1	1.0
Median (range)	5.5 (2.8–7.5)		3.1 (2.0–5.9)		30 (29–35)		2.3 (0.7–2.8)	
Bone lesions								
1	1.5		10.4		67		0.1	
2	13.7		21.2		61		0.6	
3 (1)	28.1		68.9		159		0.4	
3 (2)	14.2		49.2		159		0.3	
4	24.6		46.3		61		0.5	
5	23.9		16.7		32		1.4	
Median (range)	19.1 (1.5–28.1)		33.8 (10.4–68.9)		64 (32–159)		0.5 (0.1–1.4)	

PET: Positron emission tomography; SPECT: Single-photon emission tomography.

SPECT/CT-based mean absorbed doses (D_{ref}) and effective half-lives (tref) are reported using CT-based segmentation and OSCGM reconstruction. Activity uptakes on CT-based lesion volumes have been RC corrected for post-therapy dosimetry results. Prediction performance is expressed as the ratio of PET/CT-based predicted doses to SPECT/CT-based observed doses.

PET-predicted and SPECT-observed absorbed doses for kidneys was 2.3 (range 0.7–2.8). For lesions, the corresponding median ratio was 0.5 (range 0.1–1.4).

Using the literature effective half-lives [24], the median ratio for kidneys was 2.8 (range 0.8–3.6) and for lesions 0.6 (range 0.2–1.8). Using the reference effective half-lives in PET prediction, the median ratio was 2.3 (range 0.6–3.2) for kidneys and 1.0 (range 0.2–1.5) for lesions.

Dose predictions for combinations of the three effective half-lives and the two segmentation methods are presented in detail in Supplementary Appendix D and E.

Discussion

This pilot study evaluated the feasibility of pre-treatment dosimetry from single time-point [^{18}F]F-PSMA-1007 PET/CT to predict absorbed doses in [^{177}Lu]Lu-PSMA-I&T therapy. The main preliminary finding is that PET-based dose prediction tended to overestimate kidney absorbed dose and underestimate the bone lesion absorbed dose, with substantial interpatient and inter-lesion variability. The segmentation methods produced more consistent dosimetry results for kidneys, while greater variability was observed in lesions. The median absorbed doses were comparable to those reported in previous studies [14, 23–28].

Clearly, our pilot approach had limitations. First, the study had a limited patient cohort, and consequently such a limited sample size has no statistical power. The number of analyzed lesions per patient was limited due to the selection criteria, leading to potential high uncertainty. Further studies with a larger sample size are needed.

Furthermore, only the low-dose, noncontrast CT scans were utilized with PET and SPECT imaging. Using diagnostic CT and MRI scans could improve the accuracy of lesion delineation. The volumes of the lesions analyzed were below 21 mL, and the partial volume effect limits accurate activity estimation. Basic RC correction was carried out for post-therapy SPECT/CT scans with diligent registration to counterbalance these effects. PET-based lesion dose predictions could be improved by applying similar corrections. Despite EARL2 reconstructions, the use of two different PET/CT systems may affect the quantitation.

The SPECT-observed effective half-lives for the kidneys aligned more closely with the first set of literature-reported effective half-lives [23] than with the second set [24]. On the other hand, the SPECT-observed mean effective half-lives for bone lesions resembled the second set of literature values [24] more than the first [23]. Especially the SPECT-observed bone lesion effective half-lives exhibited a wide range of values, speaking for the need of using patient- or tumor-specific effective half-lives to achieve better accuracy in dose prediction. The accuracy of post-treatment dosimetry results could be improved using SPECT/CT scans from more time points. The results could also improve, for instance, with population pharmacokinetic modeling [29].

An avenue to explore could be to find systematic correction factors that can be applied to reduce the bias between predicted and calculated absorbed doses. This concept was researched by Peters et al. [30] using pre-treatment ^{68}Ga-PSMA-11 PET imaging to predict the absorbed dose in [^{177}Lu]Lu-PSMA-617 treatment. They reported a scaling factor of 2.2 for kidneys. Lesions showed large interpatient variation. We could not determine such tentative PET/SPECT correction factors in this pilot study.

The predictive power of pre-treatment PET imaging has been studied more extensively in the case of somatostatin receptor radionuclide therapy with statistically significant yet modest

results [31]. The kinetic difference between the short-lived PET tracers and the therapy agents ([^{18}F]F-PSMA-1007 vs. [^{177}Lu]Lu-PSMA-I&T [32]) may present a limitation that cannot be fully resolved. The short physical half-life of ^{18}F limits the time-point of PET/CT imaging, giving an image of a relatively early distribution and clearance. A potential solution to address this issue could be using a longer-lived PET or SPECT-compatible isotope to enable imaging at later time-points [33, 34].

Conclusion

In this feasibility study, dose prediction for the first [^{177}Lu]Lu-PSMA-I&T treatment cycle based on [^{18}F]F-PSMA-1007 PET/CT overestimated kidney absorbed dose and underestimated bone lesion absorbed dose compared to post-treatment SPECT/CT-based absorbed doses. Observed substantial interpatient variation suggests caution in PET-based dose prediction and the need for further studies.

Acknowledgments

The authors declare no funding was received for this work.

NACP 2026 was financially supported by the Acta Oncologica Foundation.

Conflicts of interest

VR has received institutional funding for clinical phase-I studies by a sponsor (Bayer) and a dosimetry trial (Pharmtrace). EH receives institutional funding for dosimetry services as part of a clinical trial (Pharmtrace).

Data availability statement

Research data are stored in an institutional repository and will be shared upon reasonable request to the corresponding author.

Ethics declarations & trial registry information

This study was conducted in accordance with the Declaration of Helsinki and its amendments. The study protocol was approved by the Institutional Review Board of Helsinki University Hospital (HUS/486/2025 and HUS/111/2026). As this was a retrospective study, the Institutional Review Board waived the requirement for patient informed consent, and no additional ethics statement was necessary.

Author contributions

This study bases on an extension of SK's Master's thesis, which was supervised by VR and EH, and reviewed by SS. VR designed and managed the study, including the administration of the research permission and acquiring patient data with the support of VA and MT. SPECT/CT phantom data were prepared by VR. SK performed all the quantitative analyses, result handling, and drafted the manuscript. JS and OS participated in finalizing the draft. All authors reviewed, edited, and approved the final manuscript.

References

[1] Kratochwil C, Fendler WP, Eiber M, Hofman MS, Emmett L, Calais J, et al. Joint EANM/SNMMI procedure guideline for the use of 177Lu-labeled PSMA-targeted radioligand-therapy (177Lu-PSMA-RLT). Eur J Nucl Med Mol Imaging. 2023;50(9):2830–45. https://doi.org/10.1007/s00259-023-06255-8

[2] Working group set up by the Finnish Medical Society Duodecim and the Finnish Urology Association. Prostate Cancer. Current care guidelines. Helsinki: The Finnish Medical Society Duodecim; 2025 [cited 2026 Jan 28]. Available from: www.kaypahoito.fi

[3] Castellucci P, Deandreis D, Mirzaei S, Prior J, Sattler B, Schütze C. European Nuclear Medicine Guide: a joint publication by EANM and UEMS/EBNM. European Association of Nuclear Medicine EANM, 2026 edition. Vienna. 2026.

[4] Stabin M. The practice of internal dosimetry in nuclear medicine. 1st ed. Boca Raton, FL: CRC Press; 2016.

[5] Danieli R, Milano A, Gallo S, Veronese I, Lascialfari A, Indovinaet L, et al. Personalized dosimetry in targeted radiation therapy: a look to methods, tools and critical aspects. J Pers Med. 2022;12(2):205. https://doi.org/10.3390/jpm12020205

[6] Sjögreen Gleisner K, Chouin N, Gabina PM, Cicone F, Gnesin S, Stokke C, et al. EANM dosimetry committee recommendations for dosimetry of 177Lu-labelled somatostatin-receptor- and PSMA-targeting ligands. Eur J Nucl Med Mol Imaging. 2022;49(6):1778–809. https://doi.org/10.1007/s00259-022-05727-7

[7] Ells Z, Grogan TR, Czernin J, Dahlbom M, Calais J. Dosimetry of [177Lu]Lu-PSMA-targeted radiopharmaceutical therapies in patients with prostate cancer: a comparative systematic review and meta-analysis. J Nucl Med. 2024;65(8):1264–71. https://doi.org/10.2967/jnumed.124.267452

[8] Tran-Gia J, Denis-Bacelar AM, Ferreira KM, Robinson AP, Calvert N, Fenwick AJ, et al. A multicentre and multi-national evaluation of the accuracy of quantitative Lu-177 SPECT/CT imaging performed within the MRTDosimetry project. EJNMMI Phys. 2021;8(1):55. https://doi.org/10.1186/s40658-021-00397-0

[9] Farolfi A, Armstrong WR, Djaileb L, Gafita A, Hotta M, Allen-Auerbach M, et al. Differences and common ground in 177Lu-PSMA radioligand therapy practice patterns: international survey of 95 theranostic centers. J Nucl Med. 2024;65(3):438–45. https://doi.org/10.2967/jnumed.123.266391

[10] Dickson J, Claeys W, Denis-bacelar A, Hristova I, Lassmann M, Koole M, et al. An EANM/EARL accreditation standard for quantitative Lu-177 SPECT-CT imaging. J Nucl Med. 2025;66(Supplement 1):251783–251783.

[11] Dadgar H, Karimzadeh R, Sharifi S, Hong X, Al balooshi B, Arabi H, et al. Predicting post-treatment outcomes of Lu-177 PSMA therapy using deep learning on pre-treatment Ga-68 PSMA PET images. J Nucl Med. 2025;66(Supplement 1):252132– 252132.

[12] Yazdani E, Neizehbaz A, Karamzade-Ziarati N, Emami F, Vosoughi H, Asadi M. Transforming [177Lu]Lu-PSMA-617 treatment planning: Machine learning-based radiodosiomics and swin UNETR using pretherapy PSMA positron emission tomography/computed tomography (PET/CT). Med Phys. 2025;52(10):e70030. https://doi.org/10.1002/mp.70030

[13] Yazdani E, Sadeghi M, Karamzade-Ziarati N, Jabari P, Amini P, Vosoughi H. Machine learning-based dose prediction in [177Lu]Lu-PSMA-617 Therapy by integrating biomarkers and radiomic features from [68Ga]Ga-PSMA-11 Positron emission tomography/computed tomography. Int J Radiat Oncol Biol Phys. 2025;123(3):891–908. https://doi.org/10.1016/j.ijrobp.2025.05.014

[14] Hohberg M, Reifegerst M, Drzezga A, Wild M, Schmidt M. Prediction of response to 177Lu-PSMA therapy based on tumor-to-kidney ratio on pretherapeutic PSMA PET/CT and posttherapeutic tumor-dose evaluation in mCRPC. J Nucl Med. 2023;64(11):1758–64. https://doi.org/10.2967/jnumed.122.264953

[15] Violet J, Jackson P, Ferdinandus J, Sandhu S, Akhurst T, Iravani A, et al. Dosimetry of 177Lu-PSMA-617 in metastatic castration-resistant prostate cancer: correlations between pretherapeutic imaging and whole-body tumor dosimetry with treatment outcomes. J Nucl Med. 2019;60(4):517–23. https://doi.org/10.2967/jnumed.118.219352

[16] Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. Eur J Nucl Med Mol Imaging 2017;44(Suppl 1):17–31. https://doi.org/10.1007/s00259-017-3740-2

[17] Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans Med Imaging. 1994;13(4):601–9. https://doi.org/10.1109/42.363108

[18] Wan H. Automated contouring using neural networks. White Paper TD-650. Cleveland, OH: MIM Software Inc.; 2024.

[19] Nelson AD, Nelson AS. MIM for therapy response evaluation. White Paper MIMTD-486. Cleveland, OH: MIM Software Inc.; 2018.

[20] International Atomic Energy Agency (IAEA). Quality assurance for PET and PET/CT systems. IAEA Human Health Series No. 1. Vienna: International Atomic Energy Agency; 2009.

[21] Reijonen V, Kokkonen S, Tenhunen M. Recovery coefficients in Lu-177-SPECT-CT are reduced as a function of activity concentration. EANM Innovation. 2025;1(Suppl):44–5. https://doi.org/10.1016/S3051-2913(25)00004-7

[22] Madsen MT, Menda Y, O'Dorisio TM, O'Dorisio MS. Technical note: single time point dose estimate for exponential clearance. Med Phys. 2018;45(5):2318–24. https://doi.org/10.1002/mp.12886

[23] Schuchardt C, Zhang J, Kulkarni HR, Chen X, Müller D, Baum RP. Prostate-specific membrane antigen radioligand therapy using 177Lu-PSMA I&T and 177Lu-PSMA-617 in patients with metastatic castration-resistant prostate cancer: comparison of safety, biodistribution, and dosimetry. J Nucl Med. 2022;63(8):1199–207. https://doi.org/10.2967/jnumed.121.262713

[24] Karimzadeh A, Schatz L, Sauer M, Apostolova I, Buchert R, Klutmann S, et al. Organ and tumor dosimetry including method simplification for [177Lu]Lu-PSMA-I&T for treatment of metastatic castration resistant prostate cancer. EJNMMI Phys. 2024;11(1):63. https://doi.org/10.1186/s40658-024-00668-6

[25] Okamoto S, Thieme A, Allmann J, D'Alessandria C, Maurer T, Retz M, et al. Radiation dosimetry for 177Lu-PSMA I&T in metastatic castration-resistant prostate cancer: absorbed dose in normal organs and tumor lesions. J Nucl Med. 2017;58(3):445–50. https://doi.org/10.2967/jnumed.116.178483

[26] Feuerecker B, Chantadisai M, Allmann A, Tauber R, Allmann J, Steinhelfer L, et al. Pretherapeutic comparative dosimetry of 177Lu-rhPSMA-7.3 and 177Lu-PSMA I&T in patients with metastatic castration-resistant prostate cancer. J Nucl Med. 2022;63(6):833–9. https://doi.org/10.2967/jnumed.121.262671

[27] Barna S, Haug AR, Hartenbach M, Rasul S, Grubmüller B, Kramer G, et al. Dose calculations and dose-effect relationships in 177Lu-PSMA I&T radionuclide therapy for metastatic castration-resistant prostate cancer. Clin Nucl Med. 2020;45(9):661–7. https://doi.org/10.1097/RLU.0000000000003157

[28] Baum RP, Kulkarni HR, Schuchardt C, Singh A, Wirtz M, Wiessalla S, et al. 177Lu-labeled prostate-specific membrane antigen radioligand therapy of metastatic castration-resistant prostate cancer: safety and efficacy. J Nucl Med. 2016;57(7):1006–13. https://doi.org/10.2967/jnumed.115.168443

[29] Hardiansyah D, Patrianesha BB, Glatting G. Population pharmacokinetic modeling in radiopharmaceutical therapy: a review. Front Nucl Med. 2025;5:1695332. https://doi.org/10.3389/fnume.2025.1695332

[30] Peters SMB, Hofferber R, Privé BM, de Bakker M, Gotthardt M, Janssen M, et al. [68Ga]Ga-PSMA-11 PET imaging as a predictor for absorbed doses in organs at risk and small lesions in [177Lu]Lu-PSMA-617 treatment. Eur J Nucl Med Mol Imaging. 2022;49(4):1101–12. https://doi.org/10.1007/s00259-021-05538-2

[31] Akhavanallaf A, Lu Z, Peterson AB, Blakkisrud J, Kurkowska S, Yadav S, et al. Can 177Lu-DOTATATE kidney absorbed doses be predicted from pretherapy SSTR PET? Findings from multicenter data. J Nucl Med. 2025;66(7):1082–90. https://doi.org/10.2967/jnumed.124.269098

[32] Capasso G, Stefanucci A, Tolomeo A. A systematic review on the current status of PSMA-targeted imaging and radioligand therapy. Eur J Med Chem. 2024;263:115966. https://doi.org/10.1016/j.ejmech.2023.115966

[33] Kabasakal L, AbuQbeitah M, Aygün A,Yeyin N, Ocak M, Demirci E, et al. Pre-therapeutic dosimetry of normal organs and tissues of (177) Lu-PSMA-617 prostate-specific membrane antigen (PSMA) inhibitor in patients with castration-resistant prostate cancer. Eur J Nucl Med Mol Imaging. 2015;42(13):1976–83. https://doi.org/10.1007/s00259-015-3125-3

[34] Privé BM, Derks YHW, Rosar F, Franssen GM, Peters SMB, Khreish F, et al. 89Zr-labeled PSMA ligands for pharmacokinetic PET imaging and dosimetry of PSMA-617 and PSMA-I&T: a preclinical evaluation and first in man. Eur J Nucl Med Mol Imaging. 2022;49(6):2064–76. https://doi.org/10.1007/s00259-021-05661-0