

Anthropomorphic phantom dosimetry and validation of 150–200 MeV very high-energy electron radiotherapy at CERN

Corresponding Author: Dr Fabio D'Andrea

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The study reported is interesting and important, as it provides missing high-energy benchmarks and a validated simulation framework to inform treatment-planning studies and accelerator design, advancing VHEE towards clinical application. The following aspects should be completed before it can be published:

1. There are previous works about VHEE dosimetry performed at CLEAR. The authors should give a short summary to overview these studies and highlight the advancement of the current study.
2. The validation has been shown at energies of 150 and 200 MeV. Will it still be validated at higher energy, e.g., 250 MeV, or even higher energy of UHEE, concerning the measurement and simulation method?
3. Scattered beam was produced by inserting a 200 μm -thick YAG:Ce screen, so the resulted beam size was fixed? Is it possible to vary the output beam size?
4. In Fig.2, especially Fig.2b, why the dose distributions show a varying size as the charge increases, between 4.7 and 5.1 mm?
5. Further clarifications or explanations are needed for the 1% dose difference achieved in the most cases of gamma analysis, considering the combined uncertainty of 3-5%.
6. In Table1, in the case of Run5, the gamma passing is relatively lower, and the median dose difference is larger than 1%. As discussed, this discrepancy is attributed to accumulated experimental geometric uncertainties (e.g., sequential positioning errors, minor beam jitter). So, what is the reasonable stability requirement for the future dedicated clinical machines to solve this problem?
7. As described, the experimental geometry was explicitly modelled in GATE including all the involved windows in the beam line. Is there any significant energy loss of the beam in these windows? It's better to be clarified.

Reviewer #2

(Remarks to the Author)

This manuscript presents dosimetric characterisation of Very High-Energy Electron (VHEE) beams with 150–200 MeV delivered at the CERN CLEAR facility, and the experimental validation of a GATE/Geant4 Monte Carlo (MC) framework for beam simulation and dosimetry calculations. The validation was performed with measurements in air, water, and an anatomically heterogeneous head-and-neck phantom (MARVIN), at UHDR ($>10^8$ Gy/s instantaneous dose rate) using EBT-4 film dosimetry. The gamma analysis was done using 1%/1 mm gamma criteria with 20% low-dose threshold. The work is timely, highly relevant to the re-emerging interest in VHEE and FLASH radiotherapy, and addresses a longstanding gap in experimental benchmarks for this energy range. This study represents the first detailed experimental validation of VHEE beams in the clinically critical 150–200 MeV range within a heterogeneous phantom environment, addressing a major knowledge gap. The experimental design is robust, and the agreement between MC and film is impressive. Demonstration that UHDR conditions do not break the film-MC agreement is of significant interest. However, some aspects require clarification before publication. Overall, the paper is potentially suitable for publication after revisions.

General Comments

Improper citation usage.

Lines 481–482 The authors justify using the MARVIN head-and-neck phantom to validate dose calculations in complex, heterogeneous geometries "for various radiation types, including VHEE," and they cite [28] to support MARVIN's applicability beyond photon QA to VHEE studies. The citation usage is incorrect and does not support the claim. The Ref [28] is the paper by Subiel et al., *Phys Med Biol.* 2014;59:5811. This prior work combines EBT film measurements with MC for VHEE (165 MeV), providing methodological precedent for VHEE dosimetry: another work, ref[8], by Bazalova et al. *Medical Physics*, 42(4), 2015, provides dosimetry with EBT film for 50 and 70 MeV. In both cases, the film was expocalibrate using high-energy electron beams from clinical medical accelerators. Both references [28] and [8] show that the film response is independent of energy and dose rate. Both references are directly relevant to the authors' dosimetric methodology and should be properly cited as prior work. Moreover, the results should align with the discussions in references [8] and [28].

Film dosimetry calibration using protons

The calibration was performed using 200 MeV protons, not electrons. There are multiple publications on the dependence of the EBT film response on LET. The LET spectra or VHE have a median of about 0.37, a sharp peak at 0.4 keV/um, and a shallow tail to 0.6 keV/um. Subiel et al., *Phys. Med. Biol.* 59 (2014) 5811, while proton LET spectra for 100-200 MeV are broad from 0.1 keV/um to 10 KeV/um with a median at 1 keV/um DeCunha et. al. *Med Phys.* 2025;52:3471–3480.

- The rationale for using protons for film calibration, along with potential uncertainties, should be explicitly discussed.
- Because VHEE electrons have lower LET than protons, the authors should explain why this calibration is valid for electrons at 150–200 MeV.
- Discuss potential LET-dependent response shifts and cite established EBT-4 behaviour. Discuss potential LET-related systematics for the red vs green channels in this dose range (you already justified the exclusion of blue; extend the logic to electrons vs protons).
- Provide any cross-validation (if available) from an electron beam.
- The Supplement cites AAPM TG 235 best practice language in S.1/S.3 but does not give a full bibliographic entry; please add a formal reference for TG 235 and explicitly align your protocol choices (24 h wait, 48 bit scanning, flat fielding) with that guidance.

Repeatability of the results

The manuscript does not report the lot number of the Gafchromic EBT-4 films used for calibration and measurement. This information is essential for reproducibility, as lot-to-lot variability in EBT-series films is well documented and can influence calibration curves, channel-dependent response, and final uncertainty estimation.

Film simulation

The authors do not model the EBT-4 film in the Monte Carlo geometry. The MC provides dose-to-medium, interpolated onto the film plane, while the film provides dose derived from optical density. Film–beam interactions, material composition, density differences, and energy dependence of the active layer are not simulated. This differs from prior VHEE dosimetry work (e.g., Subiel et al., 2014), where film layers were explicitly modeled. The authors should estimate the error introduced by assuming dose-to-medium $\approx$ dose-to-film, and quantify the expected impact on dose profiles and gamma pass rates (particularly in steep-gradient regions).

Beam source optics: empirical hyperbola vs direct phase space

Authors fit $\sigma(z)$ using a Fermi–Eyges–motivated hyperbola (Eq. 3; A, B, C in SI Table 3), then map to a Gaussian positional/angular source at the exit window. This is elegant and practical, but it implicitly absorbs multiple Coulomb scattering in air and any position–angle correlations into phenomenological parameters. (Since the real beam consists of the focusing and scattered components). Please provide a sensitivity test: perturb (A, B, C) within their uncertainties (SI Table 3) and show the resulting $\Delta\gamma$ pass rate and $\Delta\text{peak dose}$ in water and MARVIN—especially the seven-spot transverse ABS case—so readers can see how source model uncertainty maps to clinically relevant metrics.

Interpretation of multi-spot failure in ABS: geometry vs. physics

The reduced gamma pass rate in the ABS transverse plane (Run 5: 81.9% vs. foam: 96.97%) is mainly attributed to geometric uncertainties (pp. 7–8). However, this may also indicate limitations in the scattering model or beam-characterisation fidelity. Readers will ask whether MC lateral-scatter modelling or source ellipticity plays a role (you note subtle ellipticity in single-spot data). Please,

- Quantify the expected geometric error budget (robotic stage repeatability, phantom backlash, beam pointing jitter per spot) and simulate its convolution with the dose kernel to show whether the observed γ failure band width is reproduced.
- Explore whether a slightly anisotropic source ($\sigma_x \neq \sigma_y$ or correlated angles) consistent with your measured ellipticity (p.6, lines 209–214) changes agreement in the seven-spot ABS case.
- Comment on whether the physics list (chosen for single spot accuracy) could subtly bias cumulative multi-spot lateral tails in higher scatter media; even a statement that you tested an alternative EM option with negligible impact would help.

Insufficient uncertainty propagation

Several uncertainty sources are described qualitatively but not quantitatively propagated (film, alignment, charge stability, MC statistics).

FLASH terminology caution

The authors correctly avoid claiming a biological effect of FLASH, but the text could more explicitly state that UHDR delivery here refers only to instantaneous dose rates; biological implications cannot be drawn from dosimetry alone. The authors correctly caution that instantaneous UHDR (per spot) is not a proxy for biological FLASH outcomes. Given the active debate [5–7], please foreground this distinction earlier (Abstract or end of Introduction) and reference CLEAR-based FLASH instrumentation papers to emphasise that active, online dosimetry remains a translational bottleneck

Specific Comments

- Note: There is an inconsistency between the main text and the Supplementary Materials. SI indicates 0.5 mm³ voxels (Table S4 voxel size); the main text says 1 mm³.
- Abstract p.1, lines 14–18. 1%/1 mm γ with 20% threshold is a high bar; add a sentence justifying why this is clinically

meaningful in VHEE contexts or it is specific for the spot size (e.g., strong lateral gradients in multi-spot fields) and cite your own Gamma pass rates for single vs seven spot cases. The typical clinical rate for Gamma-analysis is 3%/3 mm, with the modern trend moving toward 2%/2mm. The clinically used low-dose threshold is 10%. Expand this explanation in the main text of the manuscript.

- Results p.4, lines 126–134. The stated 2–3% PDD agreement—please show uncertainty bands on MC and film to frame this number (and clarify whether global normalisation or local).
- Results p.6, lines 209–214. Provide σ_x/σ_y values and note the beam steering axis (superior–inferior) alignment relative to raster direction to contextualise seven-spot outcomes.
- Discussion p.9–10, lines 311–320, 352–357. Move FLASH cautionary language forward (Intro/Abstract) and cite [5–7] alongside your CLEAR dosimetry/monitoring refs [19–20] to stress that online, active dosimetry is pivotal to translation.
- Methods p.11–12, lines 380–413. Beamline description is strong; consider a compact table or inset mapping diagnostic station positions to z relative to isocentre for quick reference (complements Fig. 6).
- Suppl. Fig. S1 (p.1). Add error bars and, if possible, a film vs MC difference panel for $\sigma(z)$ and PDD at 150 MeV to mirror the main text 200 MeV presentation.
- Suppl. Table 4 (p.2). You list $0.5 \times 0.5 \times 0.5$ mm³ voxel size and $\leq 0.2\%$ target uncertainty in 90–100% isodose voxels; reconcile with $1 \times 1 \times 1$ mm³ in Methods and justify the chosen voxel size for steep gradient regions (multi-spot overlap).
- S.1 / S.3 (p.3). Remove the duplicated header and add a formal reference entry for AAPM TG 235; briefly quantify the scanner flat-field residual post-correction.

Reviewer #3

(Remarks to the Author)

The major claims of the paper are two-fold: Compared to previous works, this paper has conducted a more thorough dose measurement of VHEE beams at clinically relevant energies (150-200 MeV) and in a more realistic medium (heterogeneous phantom). It has also created a Monte Carlo simulation which has been benchmarked against these more realistic measurements.

This paper will be of interest to others in the field, as a more thorough dose measurement of VHEE beams is important, especially in verifying that Monte Carlo simulation reflects physical reality. This is an important step toward more realistic VHEE dose measurements. One of the key insights is that the experimental setup uncertainty (such as stage positioning error and beam jitter) contribute to gamma index test failure in multi-spot irradiation. The paper would likely be more influential if those arguments were quantified, as they could then be used to guide the future system design as noted by the authors.

The other point of note is the use of small beam spot sizes (which remain small even after scattering), remaining comparable in both size and energy to the earlier work by Lagzda (<https://doi.org/10.1016/j.nimb.2020.09.008>). Although the current work has included the use of an anthropomorphic phantom, and a rudimentary spot scanning experiment, it remains true that the spot size is small compared to anatomical ‘features’ of the phantom, so the results at first glance are not so dissimilar from the simpler water phantom setup. Further details are given below.

Overall, the claims are well justified for the most part, but further MC simulation is required in investigating the effect of position uncertainty and beam jitter to the deviation of dose distribution. The paper would also be more convincing with more extensive justification of statistical analysis, the particular use of the gamma index test is unusual, and a satisfying explanation is not provided. Further detailed comments follow below.

Other experiments which would strengthen the paper further include the use of a large enough beam to truly see the impact of the anthropomorphic phantom. This would showcase how much the actual patient geometry affects the dose distribution. Perhaps additional materials as a scatterer to blow up the beam would achieve this? (Detail below)

Finally, the statement that there is a “lack of comprehensive experimental data, particularly for energies above 100 MeV” implies a total absence, when in reality several works in the literature (including the one cited by the article) have been published containing at least some dosimetry aspect of VHEE beams although not specifically using the anthropomorphic phantom (<https://doi.org/10.1038/s42005-021-00536-0>, <https://doi.org/10.1088/2057-1976/abcae5>).

1. In Section 2, Results, paragraph 2, instead of describing divergence as beam spot size at arbitrary locations, it might be better to define a formula so it can be easily compared between beams. For example, defining divergence as beam spot size increase per depth increase so that 200 MeV pristine beams will have 2 mm per m divergence while 200 MeV scattered beams will have 8 mm per m divergence.

2. There are Run 1 and Run 2, which exist in Supplemental Table 2, but I can’t find where they are discussed in the main article.

3. In Fig. 3, the 3D render of the phantom is blurry and the way the slice views are presented is confusing. It’s not at all clear that Fig. 3(b)i is the axial slice of the whole phantom head, while Fig. 3 (c)i and Fig. 3(d)i are (presumably) the coronal slice of the back part of the phantom head. This figure could be made clearer by being consistent in where the slice views are located with respect to the phantom or world/beam axis.

4. It would be better to use the term axial plane instead of transverse plane, since transverse can also mean the coronal plane. This term was not used consistently throughout the article. “Transverse” is used in caption of Fig. 3 and Fig. 4, but the

dose profile uses the term “axial”. The use of Superior-to-Inferior term in Fig. 4 and Fig. 5 is also confusing, as it usually refers to top to bottom (the y-axis) instead of left to right (the x-axis). Adding a consistent world axis would also help a lot so all the plots are consistent throughout the article.

5. The use of an anthropomorphic phantom and stating that it is “an anatomically realistic, heterogeneous environment” is not entirely justified. The beam spot is very small (diameter of 1 cm), so from the beam’s eye view, the anthropomorphic phantom is not very different from a homogeneous water phantom with a slab of low-density insert in front of it. The use of an anthropomorphic phantom to showcase more realistic conditions is only valid when there are multiple beams at different angles with large field sizes (on the order of 10x10 cm), where the contour of the phantom becomes significantly different from that of a conventional water phantom.

6. The gamma index test is described in TG-218 (<https://doi.org/10.1002/mp.12810>) and most clinics (at least at the time of the publication) uses 3%/3mm with $\leq 10\%$ dose threshold. Thus, using a 20% dose threshold seems too aggressive. From Fig. 4(c) and Fig. 5(c), we can also see that the dose difference between simulation and film at the coronal plane starts to be significant near the tail. Moreover, the effect of Multiple Coulomb Scattering (which is the main scattering in VHEE beams when going through heterogeneous material such as low-density inserts in water) can only be seen near the tail as well. Therefore, using a 20% dose threshold seems to be unnecessarily eliminating the error coming this tail.

7. Further evidence should be provided that the gamma index test failure in Fig 5(c) is due to the beam jitter and stage positioning error. In other part of the article, it was claimed that the beam is stable and the stage positioning error is <0.5 mm. It is not intuitive that this kind of error would lead to what is shown, with the MC simulation uniformly overestimating the dose profile only at the flat top (one might instead expect over and under-estimation). To demonstrate this, one would need to investigate the effect of positioning uncertainty in the MC simulation to see whether such error (beam jitter and stage positioning) leads to overestimation of dose only at the flat top.

8. The effect of scattering due to the EBT-4 film (which might not necessarily be water-equivalent) was not addressed. It might or might not explain some part of the difference between simulation and film measurement. There is also no clear justification for the choice of EBT-4 over other types of EBT film (it may simply have been the only option available at the time of the experiment).

9. The statement at Section 2 about the dose profile being narrower in the low-density insert irradiation is not clear. It might be better if there were a quantitative comparison instead of only a qualitative comparison based on the figure.

Reviewer #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Physics initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The manuscript has been modified accordingly.

Reviewer #2

(Remarks to the Author)

After reviewing the revised manuscript, I find that the authors have adequately addressed all the concerns raised in the previous round of review. The manuscript has significantly improved and is now much clearer and easier to understand.

Reviewer #3

(Remarks to the Author)

The manuscript has been considerably improved. This referee is satisfied with the responses to the previous round of comments, to the best of the authors ability, which have improved the scientific accuracy of the claims made in the paper.

Reviewer #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Physics initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Note on reference numbering: Numbered citations quoted in the response-to-reviewers document refer to the reference list in the originally submitted manuscript. Citation numbers in the revised manuscript have been updated following the addition and reordering of references.

Reviewer #1:

>> The study reported is interesting and important, as it provides missing high-energy benchmarks and a validated simulation framework to inform treatment-planning studies and accelerator design, advancing VHEE towards clinical application. The following aspects should be completed before it can be published:

We thank the reviewer for their positive overall assessment of the manuscript and for their careful, constructive comments. We appreciate their recognition of the importance of this work in providing high-energy benchmarks and a validated simulation framework for VHEE.

>> 1. There are previous works about VHEE dosimetry performed at CLEAR. The authors should give a short summary to overview these studies and highlight the advancement of the current study.

We have revised the Introduction to summarise prior CLEAR-based work more explicitly and to clarify the specific advance of the present study.

“...Foundational experiments provided initial data at lower energies (e.g., 50–70 MeV in homogeneous materials [8]) or explored effects in simplified heterogeneous geometries (~150 MeV [4]). At CLEAR, subsequent studies have also reported important VHEE dosimetric and beam-characterisation data under UHDR conditions [Poppinga et al.; McManus et al.] and have supported related radiobiological investigations [Small et al.] However, experimental validation in anatomically complex heterogeneous scenarios remains limited, particularly given the technical hurdles associated with beam production and measurement at these energies [9]. Consequently, numerous computational studies suggesting VHEE offers dosimetric advantages comparable or superior to established therapies [10, 11, 12, 13] have relied on MC models that lack comprehensive experimental validation in anatomically heterogeneous settings within this high-energy regime. The present work addresses this gap by reporting detailed dosimetric measurements and MC validation at 150–200 MeV within an anthropomorphic head-and-neck phantom.”

Ref:

Poppinga D, Kranzer R, Farabolini W, Gilardi A, Corsini R, Wyrwoll V, Looe HK, Delfs B, Gabrisch L, Poppe B. VHEE beam dosimetry at CERN Linear Electron Accelerator for Research under ultra-high dose rate conditions. *Biomedical Physics & Engineering Express*. 2020;7(1). doi:10.1088/2057-1976/abcae5.

McManus M, Romano F, Lee ND, Farabolini W, Gilardi A, Royle G, Palmans H, Subiel A, et al. The challenge of ionisation chamber dosimetry in ultra-short pulsed high dose-rate Very High Energy Electron beams. *Scientific Reports*. 2020;10:9089. doi:10.1038/s41598-020-65819-y.

Small KL, Henthorn NT, Angal-Kalinin D, Chadwick AL, Santina E, Aitkenhead A, Kirkby KJ, Smith RJ, Surman M, Jones J, Farabolini W, Corsini R, Gamba D, Gilardi A, Merchant MJ, Jones RM. Evaluating very high energy electron RBE from nanodosimetric pBR322 plasmid DNA damage. *Scientific Reports*. 2021;11:3341. doi:10.1038/s41598-021-82772-6.

>> 2. The validation has been shown at energies of 150 and 200 MeV. Will it still be validated at higher energy, e.g., 250 MeV, or even higher energy of UHEE, concerning the measurement and simulation method?

The present study validates the measurement and simulation workflow specifically for the experimentally investigated 150 and 200 MeV beams. We agree that extension to higher energies, including 250 MeV and beyond, is an important future direction. The same general methodology, including experimental beam characterisation, film-based dose measurements and Monte Carlo modelling of the experimental geometry, is expected to remain applicable at energies beyond 200 MeV, including the clinically relevant extension to 250 MeV. However, validation at these energies cannot be inferred directly from the present dataset and would require dedicated beam characterisation and experimental benchmarking. For substantially

higher, GeV-scale ultra-high-energy electron (UHEE) beams, additional validation would also be required to confirm the accuracy of the underlying Monte Carlo physics models and source description. We have clarified this limitation in the Discussion.

*“...The 150–200 MeV energy range is of therapeutically critical importance for the clinical prospects of VHEE. Simulation studies consistently identify energies within this window as necessary to achieve the penetration depths required for treating deep-seated tumours (e.g., in the pelvis), where energies below 100 MeV may prove insufficient [2, 12, 13]. Consequently, numerous influential in silico investigations exploring the potential advantages of VHEE, often comparing it favourably with established photon (Volumetric Modulated Arc Therapy, VMAT) or proton (Intensity-Modulated Proton Therapy, IMPT) therapies [11, 12, 13], have operated within or encompassed this specific energy regime. **Accordingly, the present validation should be interpreted as specific to the experimentally investigated 150 and 200 MeV beams. The measurement and simulation methodology should remain applicable at energies beyond 200 MeV, including the clinically relevant extension to 250 MeV, but would require dedicated beam characterisation and experimental benchmarking under those conditions. For substantially higher, GeV-scale ultra-high-energy electron (UHEE) beams, further validation would also be needed to confirm the accuracy of the underlying Monte Carlo physics models and source description.** This study therefore substantially increases confidence in the interpretation of previous theoretical findings and...”*

>> 3. Scattered beam was produced by inserting a 200 µm-thick YAG:Ce screen, so the resulted beam size was fixed? Is it possible to vary the output beam size?

The reviewer is correct that, for a given incident beam energy, use of the 200 µm YAG:Ce screen produced a specific scattered-beam configuration in the present study. In principle, the output beam size could be varied through different scatterers and/or upstream beam tuning. However, here we deliberately used a single fixed configuration to provide a stable and reproducible basis for dosimetric benchmarking and Monte Carlo validation across the different measurement geometries. We have clarified this point in the Methods.

*“...Scattered beam: produced by inserting a 200 µm-thick YAG:Ce screen at the BTV910 diagnostic station (Fig. 6). This screen, primarily used for diagnostics, served here as a simple scattering element to deliberately broaden the beam profile, resulting in $\sigma \approx 4.5$ mm (200 MeV) or 5.5 mm (150 MeV) in air at ~52 cm. **For the present study this provided a single fixed scattered-beam configuration, although in principle the output beam size could be varied by altering the upstream scattering arrangement and beam optics.** This configuration generated beam dimensions more suitable for dosimetric characterisation across larger phantom regions and provided a test case with increased divergence...”*

>> 4. In Fig.2, especially Fig.2b, why the dose distributions show a varying size as the charge increases, between 4.7 and 5.1 mm?

The apparent variation in beam size in Fig. 2b is not interpreted as a charge-dependent change in beam optics. Quantitatively, the fitted Gaussian widths remain consistent overall, with the small variation between approximately 4.7 and 5.1 mm attributable to normal experimental and fitting variability across separate film measurements, rather than to a systematic physical change in the beam profile.

*“...Conversely, the scattered beam exhibited a significantly broader lateral dimension ($\sigma \approx 4.9$ mm) (Fig. 2b). **The small variation in fitted σ between individual irradiations was not systematic with charge and is attributed to normal experimental and fitting variability across separate film measurements.** Achieving comparable doses with the scattered beam required substantially higher charge (e.g., 24.59 nC delivered a 22.5 Gy peak dose)...”*

>> 5. Further clarifications or explanations are needed for the 1% dose difference achieved in the most cases of gamma analysis, considering the combined uncertainty of 3-5%.

The quoted 3–5% uncertainties relate to film calibration and overall dosimetric system performance, whereas the gamma analysis was applied to spatially registered relative dose distributions using global normalisation and a 20% low-dose threshold. Accordingly, the 1%/1 mm criterion was used here as a test of relative spatial and dosimetric agreement, rather than as a statement that the total absolute uncertainty at each point is 1%. We have clarified this distinction in the manuscript to avoid ambiguity.

Section 2.4

"The validation (Fig. 3) centred on the 200 MeV scattered beam, which is considered most relevant for potential clinical scenarios, across four distinct irradiation runs (Runs 3–6, Supplementary Table 1). Agreement was quantitatively evaluated using 2D dose maps, 1D profiles, and gamma analysis, applying ~~stringent, clinically-derived acceptance global~~ criteria of 1% dose difference (~~global normalisation~~) and 1 mm distance-to-agreement. ~~These stringent criteria were selected to test spatial and relative-dose agreement in the narrow VHEE fields studied here. A 20% lower dose threshold was applied so that the quantitative comparison focused on the primary high-dose plateau and steep penumbra. Passing these criteria indicates a high level of agreement between simulation and measurement in the thresholded field region...~~"

>> 6. In Table1, in the case of Run5, the gamma passing is relatively lower, and the median dose difference is larger than 1%. As discussed, this discrepancy is attributed to accumulated experimental geometric uncertainties (e.g., sequential positioning errors, minor beam jitter). So, what is the reasonable stability requirement for the future dedicated clinical machines to solve this problem?

We agree that this is an important translational question. However, the present study was not designed to derive a formal engineering tolerance for future clinical VHEE delivery systems. As detailed in later responses (Reviewer 2, Comment 4; Reviewer 3, Comment 7), we have revised the manuscript to avoid attributing the lower Run 5 pass rate solely to geometric delivery uncertainty.

The lower pass rate in Run 5 indicates that narrow multi-spot delivery in scattering media is sensitive to combined experimental and source-model uncertainties. Accordingly, future dedicated delivery systems will require highly reproducible spot delivery and rigorous geometric quality assurance, particularly for scanned multi-spot deliveries in regions with strong lateral dose gradients. However, deriving a quantitative stability requirement would require repeated irradiations and a dedicated uncertainty-propagation study using delivery hardware optimised for reproducibility, which was beyond the scope of the present validation dataset and the intended operating mode of CLEAR as a flexible research accelerator.

Discussion

"...field to combined experimental and source-model uncertainties. **For future clinical delivery systems, this supports the need for highly reproducible spot delivery and rigorous geometric quality assurance, particularly in scanned multi-spot fields.** Repeat seven-spot..."

>> 7. As described, the experimental geometry was explicitly modelled in GATE including all the involved windows in the beam line. Is there any significant energy loss of the beam in these windows? It's better to be clarified.

We agree that this point requires clarification. In the Monte Carlo workflow, the primary beam source was defined at the plane of the beamline exit window using parameters derived from downstream beam-characterisation measurements. The beamline exit window itself was not represented as a separate explicit material layer in the simulation geometry. Since the beam energy in the MC simulations was tuned to match the simulated PDDs to the experimental PDDs, the simulated energies at the source plane were different from the nominal 150 and 200 MeV beam energies (see table 2). This approach means that the Monte Carlo beam-model implicitly accounts for the net effect of beamline components upstream of the source-plane. Components downstream of the beamline exit window were modelled explicitly, including the Kapton entrance window of the water tank where applicable. For these thin downstream windows, the associated energy loss is expected to be small relative to the beam energy and intrinsic energy spread, although they still impact on beam transport through additional scattering. We have clarified this point in the Methods to distinguish between the source-model definition and the explicitly modelled downstream geometry.

"The experimental geometry downstream of the effective source plane was explicitly modelled in GATE. The primary electron source was defined at the plane of the beamline exit window using measured energy information and source parameters derived from downstream beam-characterisation measurements. Consequently, material effects upstream of this

source plane, including the vacuum exit window and the scattering configuration used to generate the broadened beam, were incorporated implicitly into the measured source description rather than treated as separate downstream energy-loss terms. This included passive downstream elements such as the air path to the measurement geometry and, where applicable, the 0.1 mm Kapton entrance window of the water tank. For these thin downstream components, the expected energy loss is small relative to the 150–200 MeV beam energy and intrinsic energy spread, although their contribution to beam transport through additional scattering was retained in the simulation geometry. Phantom geometries, ...,”

Reviewer #2:

>> This manuscript presents dosimetric characterisation of Very High-Energy Electron (VHEE) beams with 150–200 MeV delivered at the CERN CLEAR facility, and the experimental validation of a GATE/Geant4 Monte Carlo (MC) framework for beam simulation and dosimetry calculations. The validation was performed with measurements in air, water, and an anatomically heterogeneous head-and-neck phantom (MARVIN), at UHDR ($>10^8$ Gy/s instantaneous dose rate) using EBT-4 film dosimetry. The gamma analysis was done using 1%/1 mm gamma criteria with 20% low-dose threshold.

>> The work is timely, highly relevant to the re-emerging interest in VHEE and FLASH radiotherapy, and addresses a longstanding gap in experimental benchmarks for this energy range. This study represents the first detailed experimental validation of VHEE beams in the clinically critical 150–200 MeV range within a heterogeneous phantom environment, addressing a major knowledge gap. The experimental design is robust, and the agreement between MC and film is impressive. Demonstration that UHDR conditions do not break the film-MC agreement is of significant interest.

>> However, some aspects require clarification before publication. Overall, the paper is potentially suitable for publication after revisions.

We thank the reviewer for their positive overall assessment of the manuscript and for their careful, constructive comments. We appreciate their recognition of the relevance of this work and of the need for experimental benchmarking and Monte Carlo validation of VHEE beams in the clinically important 150–200 MeV range.

>> General Comments

>> Improper citation usage.

>> Lines 481–482 The authors justify using the MARVIN head-and-neck phantom to validate dose calculations in complex, heterogeneous geometries "for various radiation types, including VHEE," and they cite [28] to support MARVIN's applicability beyond photon QA to VHEE studies. The citation usage is incorrect and does not support the claim. The Ref [28] is the paper by Subiel et.al, Phys Med Biol. 2014;59:5811. This prior work combines EBT film measurements with MC for VHEE (165 MeV), providing methodological precedent for VHEE dosimetry: another work, ref[8], by Bazalova et al. Medical Physics, 42(4), 2015, provides dosimetry with EBT film for 50 and 70 MeV. In both cases, the film was expocalibrate using high-energy electron beams from clinical medical accelerators. Both references [28] and [8] show that the film response is independent of energy and dose rate. Both references are directly relevant to the authors' dosimetric methodology and should be properly cited as prior work. Moreover, the results should align with the discussions in references [8] and [28].

The reviewer is correct that the citation at this point was misplaced. Subiel et al. is relevant as prior methodological precedent for VHEE film dosimetry and Monte Carlo comparison, but it does not support the specific claim regarding MARVIN's applicability beyond photon QA. We have therefore corrected this citation usage. Specifically, we have removed the VHEE dosimetry citation from the MARVIN phantom description and instead cited the relevant prior VHEE film-dosimetry literature in the film-dosimetry methodology section, where it is more

appropriate. We also revised the MARVIN sentence so that it no longer implies that this prior VHEE work involved the MARVIN phantom.

Methods, Section 4.4

...The MARVIN phantom [18] was employed to characterise VHEE dose distributions within anatomically representative structures. Originally designed for quality assurance (QA) in head-and-neck photon radiotherapy, MARVIN's modular structure and tissue-simulating properties made it highly suitable for validating dose calculations in complex, heterogeneous geometries ~~for various radiation types, including VHEE [28]~~, moving beyond the limitations of simpler phantoms used in prior VHEE experiments [4,9]...

Methods, Section 4.5

... All film handling, irradiation and analysis procedures were carried out in accordance with established best-practice guidelines for radiochromic-film dosimetry. **The use of radiochromic film for quantitative VHEE dosimetry has also been supported by prior studies combining film measurements with Monte Carlo modelling at VHEE energies [8,28].**

>> 1. Film dosimetry calibration using protons

>> The calibration was performed using 200 MeV protons, not electrons. There are multiple publications on the dependence of the EBT film response on LET. The LET spectra of VHEE have a median of about 0.37, a sharp peak at 0.4 keV/um, and a shallow tail to 0.6 keV/um. Subiel et al., Phys. Med. Biol. 59 (2014) 5811, while proton LET spectra for 100-200 MeV are broad from 0.1 keV/um to 10 KeV/um with a median at 1 keV/um DeCunha et. al. Med Phys. 2025;52:3471–3480.

>> • The rationale for using protons for film calibration, along with potential uncertainties, should be explicitly discussed.

>> • Because VHEE electrons have lower LET than protons, the authors should explain why this calibration is valid for electrons at 150–200 MeV.

>> • Discuss potential LET-dependent response shifts and cite established EBT-4 behaviour. Discuss potential LET-related systematics for the red vs green channels in this dose range (you already justified the exclusion of blue; extend the logic to electrons vs protons).

>> • Provide any cross-validation (if available) from an electron beam.

>> • The Supplement cites AAPM TG 235 best practice language in S.1/S.3 but does not give a full bibliographic entry; please add a formal reference for TG 235 and explicitly align your protocol choices (24 h wait, 48 bit scanning, flat fielding) with that guidance.

We agree that the rationale for using a proton beam for calibration, and the implications of LET dependence, should be stated more clearly. The calibration was performed at 2 cm depth in the proximal entrance plateau of a 200 MeV proton beam, where dose-averaged LET-to-water is approximately 0.8 keV/μm, rather than in the high-LET Bragg-peak region, where LET is >5 keV/μm. For comparison, the dose-averaged LET-to-water at 2 cm depth for a 200 MeV electron beam is approximately 0.6 keV/μm. The LET values reported here were calculated using our Monte Carlo system.

We note that the response of EBT films has a dependence on LET for proton beams, as described by both Anderson et al. (2019) and Resch et al. (2020). They describe an approach in which the under-response, U , can be modelled by:

$$U = a \times LET_d + b$$

Where:

- $a = -0.0258 \text{ } \mu\text{m/keV}$
- $b = 1.0211$ (unitless)

At 0.8 keV/μm, the LET under-response of Gafchromic film is expected to be negligible (<1%), supporting the validity of a low-LET proton-based calibration for VHEE dosimetry.

We note that these under-response models were developed for EBT3 film. However, the differences between EBT3 and EBT4 primarily relate to the substrate rather than the active layer. The under-response model is therefore expected to remain relevant for EBT4 film, and our commissioning measurements for the use of EBT4 film in proton dosimetry showed an LET dependence that was consistent with the model. Specifically, LET dependence was only apparent in the Bragg-peak region, where the LET was high.

References:

Anderson SE, Grams MP, Tseung HWC, Furutani KM, Beltran CJ.

A linear relationship for the LET-dependence of Gafchromic EBT3 film in spot-scanning proton therapy.

Phys. Med. Biol. 64:055015 (2019)

DOI: doi.org/10.1088/1361-6560/ab0114

Resch AF, Heyes PD, Fuchs H, Bassler N, Georg D, Palmans H.

Dose- rather than fluence-averaged LET should be used as a single-parameter descriptor of proton beam quality for radiochromic film dosimetry.

Medical Physics 47(5):2020

DOI: doi.org/10.1002/mp.14097

We also clarify that the red and green channels were used to balance sensitivity and precision across dose. In the low-LET proton regime, no difference in LET dependence is expected between the red and green channels. Consistent with this, our film commissioning measurements showed no observable LET dependence in either the red or green channel at low LET.

Since a direct same-beam electron calibration was not performed in the present study, we have added a citation to prior cross-modality work by Whitmore et al. Finally, we have added the formal TG-235 reference and aligned the supplementary text with that guidance.

The manuscript has been revised as follows:

“Methods, Section 4.5.1

...A calibration dataset was generated with films irradiated in a clinical proton-therapy beam (nominal 200 MeV spot-scanned protons delivered by a Varian ProBeam system, uniform 10×10 cm² field) at The Christie NHS Foundation Trust (Manchester, UK). Films were placed at a depth of 2 cm in plastic water, corresponding to the proximal entrance plateau of the beam, where dose deposition is relatively uniform and LET (~0.8 keV/μm) is lower than in the Bragg-peak region (~5 keV/μm, increasing rapidly with depth distal to the Bragg peak) and more representative of the LET of a 200 MeV electron beam (~0.6 keV/μm). This geometry was selected to reduce radiation-quality-dependent response shifts associated with high-LET proton regions. This calibration strategy is also consistent with prior cross-modality radiochromic-film calibration work relevant to VHEE dosimetry, including comparisons across clinical electrons, protons, and photons in a closely related CLEAR-based study [14]. Films were irradiated to known doses between 0 and 38.4 Gy...

“Methods, Section 4.5.4

...Examination of the calibration curves (Fig. 8) showed that the red and green channels have complementary strengths. The red channel exhibits higher sensitivity and lower noise at doses below roughly 8 Gy, whereas the green channel maintains superior linearity and precision above about 15 Gy, a region where the red channel begins to saturate. In the present study, this dual-channel approach was also intended to reduce sensitivity to channel-specific systematic effects. Potential LET-related response differences between calibration and measurement conditions remain a limitation ...”

“Supplementary Methods

...All procedures were designed to adhere to established best-practice guidelines for radiochromic-film dosimetry, to ensure accuracy and reproducibility [1]. Consistent with this guidance, films were scanned after at least 24 h using 48-bit RGB acquisition at 300 dpi, with scanner-specific correction applied to reduce position-dependent response variation...

“Supplementary Ref.

[1] Niroomand-Rad A, Chiu-Tsao ST, Grams MP, Lewis DF, Marinello G, Micke A, et al. Radiochromic film dosimetry: Recommendations of AAPM Radiation Therapy Committee Task Group 235. *Med Phys.* 2020;47:e1186–e1252. <https://doi.org/10.1002/mp.14497>

>> The manuscript does not report the lot number of the Gafchromic EBT-4 films used for calibration and measurement. This information is essential for reproducibility, as lot-to-lot variability in EBT-series films is well documented and can influence calibration curves, channel-dependent response, and final uncertainty estimation.

We have added the details of the lot number to the text. EBT4 was the only suitable option at the time of the experiment. EBT3 is no longer available from the supplier as it has been superseded by EBT4. Other film varieties are aimed at different applications.

“...Quantitative dosimetry relied on EBT-4 radiochromic films (Ashland Inc., Bridgewater, NJ, USA; Lot 09122301)...”

>> 2. Film simulation

>> The authors do not model the EBT-4 film in the Monte Carlo geometry. The MC provides dose-to-medium, interpolated onto the film plane, while the film provides dose derived from optical density. Film–beam interactions, material composition, density differences, and energy dependence of the active layer are not simulated. This differs from prior VHEE dosimetry work (e.g., Subiel et al., 2014), where film layers were explicitly modeled. The authors should estimate the error introduced by assuming dose-to-medium $\approx$ dose-to-film, and quantify the expected impact on dose profiles and gamma pass rates (particularly in steep-gradient regions).

We agree that, if the films had not been represented explicitly in the Monte Carlo geometry, this would have required careful consideration of dose-to-medium versus dose-to-film differences, particularly in steep-gradient regions. We have revised the Methods section to clarify this. In the MARVIN simulations, the substrate of the EBT-4 films was explicitly included in the geometry, so the comparison was not based simply on interpolation onto an idealised film plane, and any perturbations in the dose due to the substrate would be modelled by the MC simulations. The substrate was modelled as polyester, thickness 300 μm , density 1.35 g/cm^3 . The active layer (thickness $\sim 28 \mu\text{m}$) was not explicitly modelled in the MC, but its perturbing effect on the dose is expected to be negligible due to its thinness.

“Methods, Section 4.6.1

...The MARVIN model incorporated its ABS structure, interchangeable inserts (either ABS for tissue-equivalence or low-density polyurethane foam with a density of $\approx 0.05 \text{ g}/\text{cm}^3$ to simulate air cavities), internal air cavities, and the EBT-4 films (polyester, thickness 300 μm , density 1.35 g/cm^3) at the experimental measurement locations used for comparison with the film dosimetry data. Material properties were based on standard compositions or, for specific components like the foam, defined explicitly. For MARVIN simulations, AUTOMC performed the mapping of CT Hounsfield Units to material definitions (ABS, foam, air) and their respective densities. The inclusion of the film geometry was used to account for potential local perturbations to fluence and dose deposition at the film measurement planes...”

>> 3. Beam source optics: empirical hyperbola vs direct phase space

>> Authors fit $\sigma(z)$ using a Fermi–Eyges–motivated hyperbola (Eq. 3; A, B, C in SI Table 3), then map to a Gaussian positional/angular source at the exit window. This is elegant and practical, but it implicitly absorbs multiple Coulomb scattering in air and any position–angle correlations into phenomenological parameters. (Since the real beam consists of the focusing and scattered components). Please provide a sensitivity test: perturb (A, B, C) within their uncertainties (SI Table 3) and show the resulting $\Delta\gamma$ pass rate and $\Delta\text{peak dose}$ in water and MARVIN—especially the seven-spot transverse ABS case—so readers can see how source model uncertainty maps to clinically relevant metrics.

We agree that the fitted hyperbolic source model is an effective, measurement-driven description of the in-air beam evolution, rather than an explicit component-by-component reconstruction of upstream phase-space correlations and scattering contributions.

In the present study, the adequacy of this source model was assessed through agreement with independent measurements in air, water and the MARVIN phantom. The fitted parameter uncertainties reported in Supplementary Table 3 are small, and the derived source model reproduced the measured beam broadening and PDD behaviour in air and water, as well as the high agreement observed in the single-spot and sagittal MARVIN cases.

We did not perform a formal perturbation sweep of the fitted A, B and C parameters through the water and MARVIN validation simulations. These are correlated fit parameters, so perturbing them independently within their uncertainties would not necessarily represent physically meaningful variations of the source model. An analysis of inter-related variations in A,B,C is possible, but such an analysis would require a dedicated Monte Carlo uncertainty-propagation study, including assumptions about parameter covariance and the magnitude of realistic source perturbations, and is beyond the scope of the present study. Because the number of repeat MARVIN irradiations was limited, robust perturbation bounds could not be derived from those measurements alone. The MARVIN measurements were designed as validation cases rather than repeat datasets from which to derive empirical perturbation bounds.

While not a direct perturbation of A, B and C, we performed broader targeted sensitivity checks to probe the robustness of the source description, particularly for the seven-spot delivery. As detailed further in our response to Comment 4 below, these included spot-weighting variability, angular offsets of the incident beam up to 1° in 0.1° increments, source-energy and energy-spread variation. These checks did not identify a single source or transport parameter that accounted for the isolated seven-spot ABS posterior oblique plane discrepancy while preserving the high agreement observed in the other validation cases.

We have revised the manuscript to make clearer that the source model was derived from measured in-air beam evolution rather than direct upstream phase-space tracking. We have also revised the Results presentation so that Fig. 1a is identified as the source-model derivation step, while Fig. 1b,c are presented as validation of the derived model against independent measurements in air and water. We agree that a formal uncertainty propagation study would be valuable for fully quantifying this sensitivity and have noted this as a limitation of the current work.

Results, Section 2.1 heading:

2.1 In-air beam characterisation and source-model derivation

“Results, Section 2.1

...The experimental isocentre for the MARVIN phantom measurements was defined 30 cm downstream from the beam exit window (Section 4.4.1). Distances reported here are relative to this isocentre unless specified otherwise. To derive the source-model parameters used in the subsequent MC simulations (Section 4.6.3), the in-air evolution of beam spot size, σ , was measured for 150 MeV and 200 MeV electron beams under pristine (unscattered) and scattered conditions (Fig. 1a)...”

“Results, Section 2.2 heading

2.2 Monte Carlo validation in air and water”

“Results, Section 2.2

Using the source-model parameters derived from the in-air measurements in Fig. 1a (Supplementary Table 3), the MC simulation framework was validated against independent experimental data acquired in both water and air (Fig. 1b,c).”

Discussion

“...Several limitations of this study should be acknowledged. The derivation of beam optics parameters from downstream in-air measurements, whilst validated by the overall agreement, is an approximation compared with direct phase-space measurements in vacuum. Although the fitted source model was validated against independent measurements, a full propagation of source-model uncertainty through all phantom configurations was beyond the scope of this study and would require additional repeated measurements and simulations. Furthermore, the beam commissioning data used for the source model (Fig. 1a) did not extend to cover the exact MARVIN isocentre position. This represents a practical compromise necessary when working within a versatile research facility like CLEAR, but one that would be avoidable in a dedicated clinical facility...”

>> 4. Interpretation of multi-spot failure in ABS: geometry vs. physics

>> The reduced gamma pass rate in the ABS transverse plane (Run 5: 81.9% vs. foam: 96.97%) is mainly attributed to geometric uncertainties (pp. 7–8). However, this may also

indicate limitations in the scattering model or beam-characterisation fidelity. Readers will ask whether MC lateral-scatter modelling or source ellipticity plays a role (you note subtle ellipticity in single-spot data). Please,

>> • Quantify the expected geometric error budget (robotic stage repeatability, phantom backlash, beam pointing jitter per spot) and simulate its convolution with the dose kernel to show whether the observed γ failure band width is reproduced.

>> • Explore whether a slightly anisotropic source ($\sigma_x \neq \sigma_y$ or correlated angles) consistent with your measured ellipticity (p.6, lines 209–214) changes agreement in the seven-spot ABS case.

>> • Comment on whether the physics list (chosen for single spot accuracy) could subtly bias cumulative multi-spot lateral tails in higher scatter media; even a statement that you tested an alternative EM option with negligible impact would help.

We agree that the original wording attributed the reduced gamma pass rate in the Run 5 ABS posterior oblique-plane case too specifically to geometric delivery uncertainty. The present data cannot isolate a single cause, and residual source-characterisation effects, registration uncertainty, lateral-scatter modelling and experimental positioning uncertainty could all contribute.

To assess whether straightforward modelling choices could explain the discrepancy, we performed targeted sensitivity checks. These included comparing uniform spot weighting with the measured variance in spot-to-spot charge weighting. The measured charges varied modestly: 8.2 to 8.6 nC in Run 5, corresponding to 8.46 ± 0.18 nC, and 8.5 to 9.3 nC in Run 6, corresponding to 8.86 ± 0.30 nC. Incorporating these measured weights did not materially alter the calculated dose distribution or explain the Run 5 discrepancy. We also explored angular offsets from 0 to 1° in 0.1° increments, source-energy variation using monoenergetic simulations in 1 MeV increments, different energy-spread assumptions and alternative Geant4 EM physics configurations. None identified a single modelling choice that could explain the isolated Run 5 posterior oblique-plane discrepancy while preserving the high agreement observed in the single-spot cases and the corresponding sagittal seven-spot comparison.

The measured profiles indicate mild lateral anisotropy, with the experimental beam slightly broader in the superior–inferior direction than in the left–right direction, consistent with the subtle ellipticity observed in the single-spot measurements. However, this anisotropy alone does not explain the plateau-region discrepancy in Fig. 5c.

We have therefore revised the manuscript to avoid presenting source ellipticity, beam jitter or stage positioning as a definitive single explanation. The Run 5 result is now interpreted as sensitivity of the narrow multi-spot field to combined experimental and source-model uncertainties. A full uncertainty propagation would require repeat seven-spot ABS irradiations and a dedicated simulation campaign, which is now explicitly acknowledged as a limitation in the revised Discussion.

Section 2.4.2

*“In the **posterior oblique** plane (Fig. 5c,d), validation captured the cumulative lateral fluence, creating an oblong dose distribution aligned with the **seven-spot delivery direction along the left–right axis**. For the water-equivalent insert (Run 5), the MC simulation reproduced the overall flat-topped field shape, although the 2-D gamma map showed a central failure band aligned with the seven-spot delivery direction, with additional peripheral clusters (Fig. 5c). In contrast, the foam insert (Run 6) maintained visibly uniform agreement, with a measured plateau intensity of approximately 27.4 Gy closely matched by the simulation within $\pm 2\%$ (Fig. 5d). The resulting gamma pass rates were 81.9% for the water-equivalent insert and 96.97% for the foam insert (Table 1). This isolated discrepancy was not reproduced in the single-spot or sagittal seven-spot comparisons and is therefore best interpreted as sensitivity of the narrow multi-spot field to combined experimental and source-model uncertainties...”*

Discussion

*“Beyond model validation, this work revealed an important practical consideration for future multi-spot VHEE delivery. The reduced gamma pass rate for the seven-spot **posterior oblique-plane** plan within the water-equivalent MARVIN insert (Run 5, 81.9% at 1%/1 mm, Fig. 5c), when contrasted with the high agreement achieved in the less-scattering foam insert (Run 6, 96.97%,*

Fig. 5d), indicates that this configuration was particularly sensitive to small experimental or modelling uncertainties. Since high agreement was maintained in the sagittal view and for all single-spot cases, this discrepancy does not indicate a general failure of the MC framework, but rather reflects the sensitivity of a narrow overlapping multi-spot field to combined experimental and source-model uncertainties. Repeat seven-spot ABS irradiations were not available, limiting our ability to quantify experimental variability or derive a formal geometric error budget. In addition, the seven-spot delivery was produced by translating the phantom relative to a stationary beam, rather than by clinical-style magnetic scanning. These constraints are characteristic of using a versatile research accelerator such as CLEAR for an experimental approximation of spot-scanned delivery, where delivery and control systems are designed for experimental flexibility rather than the optimised reproducibility expected of a dedicated clinical scanned-beam system..."

>> 5. Insufficient uncertainty propagation

>> Several uncertainty sources are described qualitatively but not quantitatively propagated (film, alignment, charge stability, MC statistics).

Quantitative uncertainty estimates are already provided for the main contributors, including film-related uncertainty, positioning uncertainty, charge handling, and MC statistical uncertainty. We agree, however, that the summary can be made clearer. Section 4.7 now states explicitly that the quoted 3–5% uncertainty refers to absolute dose comparison and is not directly equivalent to the 1% dose-difference criterion used in the globally normalised gamma analysis.

Section 4.7 Statistical analysis

...MC simulations (kept below 0.2 % in high-dose regions; see Section 4.6.5). Shot-to-shot variations in beam charge were monitored and incorporated in the dose calculation for each measurement, minimising their impact on the relative-dose uncertainty. This uncertainty estimate applies to absolute dose comparison and should not be interpreted as directly equivalent to the 1 % dose-difference criterion used in the globally normalised gamma analysis.

>> 6. FLASH terminology caution

>> The authors correctly avoid claiming a biological effect of FLASH, but the text could more explicitly state that UHDR delivery here refers only to instantaneous dose rates; biological implications cannot be drawn from dosimetry alone. The authors correctly caution that instantaneous UHDR (per spot) is not a proxy for biological FLASH outcomes. Given the active debate [5–7], please foreground this distinction earlier (Abstract or end of Introduction) and reference CLEAR-based FLASH instrumentation papers to emphasise that active, online dosimetry remains a translational bottleneck

We thank the reviewer for recognising that the manuscript already sought to distinguish physical UHDR delivery from biological FLASH outcomes. We agree that this should be slightly clearer. Rather than adding a long qualification in the early Introduction, we have kept the background concise and revised the later wording so that dosimetric demonstration of UHDR conditions is not presented as evidence of a biological FLASH effect. We have also added a brief earlier clarification and cited the CLEAR-based active dosimetry literature.

Introduction

"...including air cavity analogues), offered a more clinically relevant and challenging test environment than the simpler geometries employed in previous VHEE experiments [4, 8]. All irradiations in this study were delivered under instantaneous UHDR conditions. This denotes the physical beam-delivery regime and does not in itself imply a biological FLASH effect. Reliable active online dosimetry for such pulsed UHDR VHEE beams also remains an important translational challenge [19,20]."

>> 7. Specific Comments

>> • Note: There is an inconsistency between the main text and the Supplementary Materials. SI indicates 0.5 mm³ voxels (Table S4 voxel size); the main text says 1 mm³.

...These resulting 3D dose distributions, scored in 0.5 × 0.5 × 0.5 mm³ voxels, were then validated against the experimental EBT-4 film measurements...

>> • Abstract p.1, lines 14–18. 1%/1 mm γ with 20% threshold is a high bar; add a sentence justifying why this is clinically meaningful in VHEE contexts or it is specific for the spot size (e.g., strong lateral gradients in multi-spot fields) and cite your own Gamma pass rates for single vs seven spot cases. The typical clinical rate for Gamma-analysis is 3%/3 mm, with the modern trend moving toward 2%/2mm. The clinically used low-dose threshold is 10%. Expand this explanation in the main text of the manuscript.

We agree that the use of global 1%/1 mm gamma criteria with a 20% low-dose threshold requires clearer justification. The comparison with more familiar clinical QA conventions is useful context, but the present study addresses a different problem: experimental validation of Monte Carlo dose predictions against film measurements for narrow VHEE spots and multi-spot fields, rather than routine clinical photon IMRT QA. In this setting, the 1%/1 mm criteria were selected to probe high-resolution spatial and relative-dose agreement, and the 20% threshold was used so that the quantitative comparison focused on the primary high-dose plateau and steep penumbra of these narrow fields.

Section 2.4

"The validation (Fig. 3) centred on the 200 MeV scattered beam, which is considered most relevant for potential clinical scenarios, across four distinct irradiation runs (Runs 3–6, Supplementary Table 1). Agreement was quantitatively evaluated using 2D dose maps, 1D profiles, and gamma analysis, applying ~~stringent, clinically-derived acceptance~~ global criteria of 1% dose difference (~~global normalisation~~) and 1 mm distance-to-agreement. ~~These stringent criteria were selected to test spatial and relative-dose agreement in the narrow VHEE fields studied here. A 20% lower dose threshold was applied so that the quantitative comparison focused on the primary high-dose plateau and steep penumbra. Passing these criteria indicates a high level of agreement between simulation and measurement in the thresholded field region...~~"

>> • Results p.4, lines 126–134. The stated 2–3% PDD agreement—please show uncertainty bands on MC and film to frame this number (and clarify whether global normalisation or local).

The PDD curves in Fig. 1c are normalised to the maximum measured dose in water, so the reported 2–3% agreement refers to relative agreement on this common globally normalised scale, not local point-by-point normalisation. We have added representative film-measurement error bars to Fig. 1 and clarified the uncertainty basis in the caption.

"...Simulated central-axis percentage depth-dose (PDD) curves agreed with measurements to within 2–3% in both water and air between +20 cm and +45 cm relative to the isocentre (Fig. 1c). All PDDs are shown relative to the maximum measured dose in water, enabling direct comparison of attenuation in water and air..."

Fig.1

"....Error bars indicate film-measurement uncertainty: $\pm 3\%$ for fitted spot-size σ and $\pm 2.5\%$ for normalised PDD..."

>> • Results p.6, lines 209–214. Provide σ_x/σ_y values and note the beam steering axis (superior–inferior) alignment relative to raster direction to contextualise seven-spot outcomes.

This point is addressed in our response to Reviewer 3, Comment 4. In brief, we have revised the Methods to define the experimental coordinate system explicitly and to clarify that the seven-spot delivery was produced by translating the phantom along the left–right axis relative to the stationary beam. Figure captions and in-text references have also been revised to follow this convention throughout.

Manuscript change:

See response to Reviewer 3, Comment 4.

>> • Discussion p.9–10, lines 311–320, 352–357. Move FLASH cautionary language forward (Intro/Abstract) and cite [5–7] alongside your CLEAR dosimetry/monitoring refs [19–20] to stress that online, active dosimetry is pivotal to translation.

This point is addressed above in Reviewer 2, Comment 6. We have moved the FLASH cautionary language earlier to clarify that UHDR refers only to the physical instantaneous dose-rate regime, and added CLEAR-based monitoring/dosimetry references to highlight active online dosimetry as a translational challenge.

*Manuscript change:
See response to Reviewer 2, Comment 6.*

>> • Methods p.11–12, lines 380–413. Beamline description is strong; consider a compact table or inset mapping diagnostic station positions to z relative to isocentre for quick reference (complements Fig. 6).

The diagnostic station positions and their relationship to the beam exit window and experimental isocentre are described in the Methods and illustrated schematically in Fig. 6. To avoid adding a further table that would duplicate this information, we have instead clarified the coordinate reference in the Methods, including the definition of the experimental isocentre and the source-plane position relative to it.

“..The experimental isocentre was defined as the point on the central beam axis 30 cm downstream of the beam exit window, and distances in the experimental area are reported relative to this position unless otherwise stated”

>> • Suppl. Fig. S1 (p.1). Add error bars and, if possible, a film vs MC difference panel for $\sigma(z)$ and PDD at 150 MeV to mirror the main text 200 MeV presentation.

We have added representative film-measurement error bars to Fig. S1 and clarified the uncertainty basis in the caption to match main text 200 MeV.

>> • Suppl. Table 4 (p.2). You list $0.5 \times 0.5 \times 0.5 \text{ mm}^3$ voxel size and $\leq 0.2\%$ target uncertainty in 90–100% isodose voxels; reconcile with $1 \times 1 \times 1 \text{ mm}^3$ in Methods and justify the chosen voxel size for steep gradient regions (multi-spot overlap).

We thank the reviewer for identifying this inconsistency. The MC dose scoring grid was $0.5 \times 0.5 \times 0.5 \text{ mm}^3$, as stated in Supplementary Table 4. The Methods text has been corrected accordingly. This voxel size was selected to preserve spatial resolution in the steep lateral gradients and multi-spot overlap regions while maintaining feasible computation time.

>> • S.1 / S.3 (p.3). Remove the duplicated header and add a formal reference entry for AAPM TG 235; briefly quantify the scanner flat-field residual post-correction.

We thank the reviewer for noting these issues. The duplicated Supplementary header has been removed, and a formal reference to AAPM TG-235 has been added to the Supplementary references. We have also clarified the flat-field correction description.

*Manuscript change:
See response to Reviewer 2, Comment 1.*

Reviewer #3:

>> The major claims of the paper are two-fold: Compared to previous works, this paper has conducted a more thorough dose measurement of VHEE beams at clinically relevant energies (150–200 MeV) and in a more realistic medium (heterogeneous phantom). It has also created

a Monte Carlo simulation which has been benchmarked against these more realistic measurements.

>> This paper will be of interest to others in the field, as a more thorough dose measurement of VHEE beams is important, especially in verifying that Monte Carlo simulation reflects physical reality. This is an important step toward more realistic VHEE dose measurements. One of the key insights is that the experimental setup uncertainty (such as stage positioning error and beam jitter) contribute to gamma index test failure in multi-spot irradiation. The paper would likely be more influential if those arguments were quantified, as they could then be used to guide the future system design as noted by the authors.

>> The other point of note is the use of small beam spot sizes (which remain small even after scattering), remaining comparable in both size and energy to the earlier work by Lagzda (<https://doi.org/10.1016/j.nimb.2020.09.008> [doi.org]). Although the current work has included the use of an anthropomorphic phantom, and a rudimentary spot scanning experiment, it remains true that the spot size is small compared to anatomical 'features' of the phantom, so the results at first glance are not so dissimilar from the simpler water phantom setup. Further details are given below.

>> Overall, the claims are well justified for the most part, but further MC simulation is required in investigating the effect of position uncertainty and beam jitter to the deviation of dose distribution. The paper would also be more convincing with more extensive justification of statistical analysis, the particular use of the gamma index test is unusual, and a satisfying explanation is not provided. Further detailed comments follow below.

>> Other experiments which would strengthen the paper further include the use of a large enough beam to truly see the impact of the anthropomorphic phantom. This would showcase how much the actual patient geometry affects the dose distribution. Perhaps additional materials as a scatterer to blow up the beam would achieve this? (Detail below)

>> Finally, the statement that there is a "lack of comprehensive experimental data, particularly for energies above 100 MeV" implies a total absence, when in reality several works in the literature (including the one cited by the article) have been published containing at least some dosimetry aspect of VHEE beams although not specifically using the anthropomorphic phantom (<https://doi.org/10.1038/s42005-021-00536-0> [doi.org], <https://doi.org/10.1088/2057-1976/abcae5> [doi.org]).

We thank the reviewer for their careful and constructive assessment of the manuscript. We appreciate their recognition of the importance of more realistic VHEE dose measurements and MC validation. In response, we have clarified the scope of the anthropomorphic phantom validation, revised the coordinate and plane terminology, strengthened the gamma-analysis rationale, and moderated the interpretation of the seven-spot discrepancy.

>> 1. In Section 2, Results, paragraph 2, instead of describing divergence as beam spot size at arbitrary locations, it might be better to define a formula so it can be easily compared between beams. For example, defining divergence as beam spot size increase per depth increase so that 200 MeV pristine beams will have 2 mm per m divergence while 200 MeV scattered beams will have 8 mm per m divergence.

We appreciate the reviewer's suggestion to express the beam broadening using a simpler comparative divergence metric. We have revised the relevant text in the paper. However, for beam parametrisation we use a hyperbolic fit, as described in the subsequent paragraph in the manuscript, to provide a more realistic characterisation of the beam optics. This is particularly relevant for uncollimated electron beams, where the spatial spread does not

evolve as a simple linear expansion from a point source, because it reflects both the initial beam phase space and multiple Coulomb scattering in air/water. We therefore consider the fitted hyperbolic model used for the MC source definition to be the more physically appropriate and informative description in this context.

>> 2. There are Run 1 and Run 2, which exist in Supplemental Table 2, but I can't find where they are discussed in the main article.

Runs 1 and 2 were included in the Supplementary Table for completeness, but the main text focused on the scattered-beam cases (Runs 3–6) because these provided the more demanding and clinically relevant validation of the Monte Carlo model, particularly for beam broadening and dose transport through heterogeneous media. In addition, the broader scattered-beam configuration produced spot sizes more representative of the scale typically considered relevant for scanned therapeutic beam delivery, making these cases the more informative validation set for the present study. To avoid confusion, we have clarified this in the manuscript.

Section 2.4

....Runs 1 and 2 are reported in the Supplementary Information for completeness; the main-text MARVIN validation focuses on the 200 MeV scattered-beam cases (Runs 3–6), which provided the most relevant heterogeneous-phantom comparisons.....

>> 3. In Fig. 3, the 3D render of the phantom is blurry and the way the slice views are presented is confusing. It's not at all clear that Fig. 3(b)i is the axial slice of the whole phantom head, while Fig. 3 (c)i and Fig. 3(d)i are (presumably) the coronal slice of the back part of the phantom head. This figure could be made clearer by being consistent in where the slice views are located with respect to the phantom or world/beam axis.

We agree that the original Fig. 3 caption and panel labelling were unclear. We have revised the figure presentation and caption to define the slice orientations more explicitly. The updated caption now distinguishes the 3D-rendered whole-phantom view from the sagittal and posterior oblique film/slice views, and indicates that the dashed lines in the sagittal panels mark the location of the posterior oblique plane shown in the corresponding panels. This should make the relationship between the 3D phantom rendering, the CT slices, the MC dose distributions and the registered film measurements clearer.

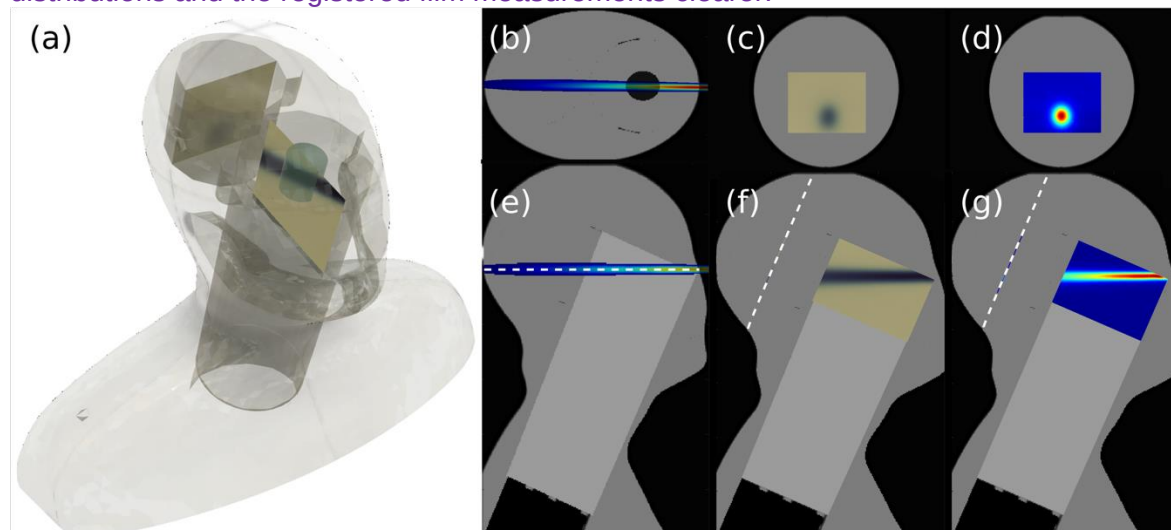


Fig. 3: Validation of the Monte Carlo (MC) simulations by experimental film dosimetry. (a) 3D rendering of the MARVIN head phantom, showing the interchangeable insert and the orientation of the measurement planes. (b) Axial CT slice through the phantom showing the MC simulated dose distribution along the beam path. (c) Posterior oblique CT slice at the film position, with the raw EBT-4 film measurement overlaid. (d) Corresponding processed and registered film dose distribution in the CT coordinate system, used for direct comparison with the MC simulation. (e) Sagittal CT slice showing the simulated dose distribution along the beam direction. (f) Sagittal CT slice with the raw EBT-4 film measurement overlaid. (g) Corresponding processed and registered sagittal film dose distribution. Dashed lines in the

sagittal panels indicate the approximate locations of the axial reference slice shown in (b) and the posterior oblique film plane shown in (c,d).

>> 4. It would be better to use the term axial plane instead of transverse plane, since transverse can also mean the coronal plane. This term was not used consistently throughout the article. “Transverse” is used in caption of Fig. 3 and Fig. 4, but the dose profile uses the term “axial”. The use of Superior-to-Inferior term in Fig. 4 and Fig. 5 is also confusing, as it usually refers to top to bottom (the y-axis) instead of left to right (the x-axis). Adding a consistent world axis would also help a lot so all the plots are consistent throughout the article.

We thank the reviewer for highlighting these inconsistencies. We agree that the terminology should be made clearer and more consistent throughout the manuscript. We have also clarified the coordinate system explicitly in the Methods and corrected the description of the ‘spot-scanning’ so that the stated translation direction is consistent with the defined axes. We have also replaced the previous inconsistent use of “transverse” and “coronal” with the final plane terminology used in the revised figures, including “posterior oblique” for the posterior film plane where appropriate. Figure captions and in-text references have been revised to follow this convention throughout.

Methods, Section 4.4.1

“The centre of the MARVIN phantom was positioned at the previously defined isocentre, 30 cm downstream of the beam exit window (Fig. 7a). **For clarity, the experimental coordinate system was defined as x = left–right, y = superior–inferior, and z = beam direction, propagating anterior-to-posterior through the phantom.** The phantom was mounted on a translation stage allowing efficient alternation between phantom irradiation and in-air beam verification measurements. Multi-spot irradiations were created by translating the phantom laterally **along the left–right axis relative** to the stationary beam between subsequent spot deliveries, to mimic a discrete spot-scanning delivery technique. A seven-spot linear pattern with 5 mm centre-to-centre spacing along the **x-axis** was delivered for selected runs (Supplementary Table 1).”

>> 5. The use of an anthropomorphic phantom and stating that it is “an anatomically realistic, heterogeneous environment” is not entirely justified. The beam spot is very small (diameter of 1 cm), so from the beam’s eye view, the anthropomorphic phantom is not very different from a homogeneous water phantom with a slab of low-density insert in front of it. The use of an anthropomorphic phantom to showcase more realistic conditions is only valid when there are multiple beams at different angles with large field sizes (on the order of 10x10 cm), where the contour of the phantom becomes significantly different from that of a conventional water phantom.

We agree that, for a narrow 200 MeV beam, the influence of the full external contour of the head-and-neck phantom is less pronounced than it would be for larger fields or multi-angle delivery. However, we consider the MARVIN phantom to remain a more representative heterogeneous test geometry than a simple slab configuration, because the beam traverses non-planar internal interfaces, including curved low-density cavities. This is particularly relevant in the multi-spot measurements, where overlapping penumbra interact with these internal heterogeneous boundaries. In addition, the use of a CT-derived phantom geometry and a dose-calculation workflow closer to radiotherapy practice makes this a more clinically relevant validation setting than a simple slab-based experiment. We have therefore moderated the wording in the manuscript to reflect this more precisely.

Section 4.4

...The MARVIN phantom [18] was employed to characterise VHEE dose distributions within anatomically representative structures. Originally designed for quality assurance (QA) in head-and-neck photon radiotherapy, MARVIN’s modular structure and tissue-simulating properties made **it well suited to validating dose calculations in complex, heterogeneous geometries, extending beyond the simpler phantoms used in prior VHEE experiments [4,9]. In the narrow-beam setting studied here, this relevance arises primarily from the CT-derived heterogeneous geometry and the curved internal density interfaces encountered by the beam, rather than from the external contour alone.** MARVIN was fabricated in discrete modular segments from acrylonitrile butadiene styrene (ABS), a plastic exhibiting tissue-equivalent behaviour for high-energy radiation interactions...

>> 6. The gamma index test is described in TG-218 (<https://doi.org/10.1002/mp.12810> [doi.org]) and most clinics (at least at the time of the publication) uses 3%/3mm with $\leq 10\%$ dose threshold. Thus, using a 20% dose threshold seems too aggressive. From Fig. 4(c) and Fig. 5(c), we can also see that the dose difference between simulation and film at the coronal plane starts to be significant near the tail. Moreover, the effect of Multiple Coulomb Scattering (which is the main scattering in VHEE beams when going through heterogeneous material such as low-density inserts in water) can only be seen near the tail as well. Therefore, using a 20% dose threshold seems to be unnecessarily eliminating the error coming this tail.

We agree that the use of a 20% low-dose threshold requires clearer justification. As also noted in response to Reviewer 2, comparison with TG-218-style clinical PSQA is useful context, but the present study addresses a different problem: experimental validation of Monte Carlo dose predictions against film measurements for narrow VHEE spots, rather than routine clinical photon IMRT QA. In this setting, the 20% threshold was chosen so that the quantitative comparison focused on the primary high-dose plateau and steep penumbra, rather than being diluted by the large number of low-dose points outside the principal field region. We agree, however, that this rationale should have been made clearer. We have therefore revised the manuscript to state more explicitly that the reported gamma results apply to the thresholded region in this experimental VHEE validation framework, rather than representing a direct analogue of standard clinical PSQA metrics.

Section 2.4

"The validation (Fig. 3) centred on the 200 MeV scattered beam, which is considered most relevant for potential clinical scenarios, across four distinct irradiation runs (Runs 3–6, Supplementary Table 1). Agreement was quantitatively evaluated using 2D dose maps, 1D profiles, and gamma analysis, applying ~~stringent, clinically derived acceptance~~ **global** criteria of 1% dose difference (~~global normalisation~~) and 1 mm distance-to-agreement. ~~These stringent criteria were selected to test spatial and relative-dose agreement in the narrow VHEE fields studied here. A 20% lower dose threshold was applied so that the quantitative comparison focused on the primary high-dose plateau and steep penumbra (Table 1).~~ Passing these criteria indicates a high level of agreement between simulation and measurement **in the thresholded field region...**"

>> 7. Further evidence should be provided that the gamma index test failure in Fig 5(c) is due to the beam jitter and stage positioning error. In other part of the article, it was claimed that the beam is stable and the stage positioning error is <0.5 mm. It is not intuitive that this kind of error would lead to what is shown, with the MC simulation uniformly overestimating the dose profile only at the flat top (one might instead expect over and under-estimation). To demonstrate this, one would need to investigate the effect of positioning uncertainty in the MC simulation to see whether such error (beam jitter and stage positioning) leads to overestimation of dose only at the flat top.

We agree with the reviewer that the original wording was too specific in attributing the Run 5 ABS postposterior oblique plane discrepancy to beam jitter and stage positioning. The observed difference in the flat-top/overlap region cannot, on its own, prove that mechanical delivery error was the dominant cause. However, in an overlapping multi-spot field, small unmodelled variations in relative spot position, effective spot size, beam pointing, registration or source description can act as an effective spatial broadening of the delivered field. This can redistribute fluence from the high-dose overlap region into the surrounding penumbra, reducing the measured plateau relative to an idealised MC delivery with fixed spot positions and source parameters. This provides a plausible explanation for why the discrepancy is most visible in the flat-top/overlap region, but we agree that it should not be presented as proof of beam jitter or stage positioning alone.

As detailed in our response to Reviewer 2 (Comment 4), we explored several alternative modelling choices, including measured spot weighting, small angular offsets, source-energy

variations, different energy-spread assumptions and an alternative Geant4 physics configuration. These checks did not identify a single parameter that uniquely explained the isolated Run 5 posterior posterior oblique plane behaviour while maintaining the high agreement observed in the other validation cases. We have therefore revised the manuscript to describe the result more cautiously as the sensitivity of the narrow multi-spot field to combined experimental and source-model uncertainties. We also acknowledge that repeat seven-spot ABS irradiations were not available, limiting our ability to quantify the experimental variability of this specific configuration.

*Manuscript change:
See response to Reviewer 2, Comment 4.*

>> 8. The effect of scattering due to the EBT-4 film (which might not necessarily be water-equivalent) was not addressed. It might or might not explain some part of the difference between simulation and film measurement. There is also no clear justification for the choice of EBT-4 over other types of EBT film (it may simply have been the only option available at the time of the experiment).

We agree that this should have been stated more clearly. As noted above in response to Reviewer 2, the EBT-4 film was not treated only as an idealised measurement plane: the film was included in the MARVIN Monte Carlo geometry at the experimental measurement locations, so potential local perturbations from the film were represented in the simulation. We have revised the Methods to clarify this point. We have also added a brief rationale for using EBT-4, namely its high spatial resolution and suitability for dosimetry in small, steep-gradient VHEE fields.

“Methods, Section 4.6.1

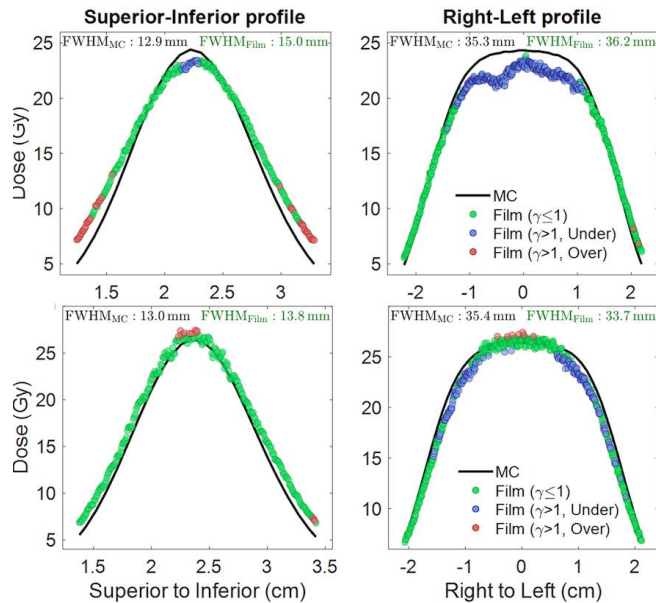
...The MARVIN model incorporated its ABS structure, interchangeable inserts (either ABS for tissue-equivalence or low-density polyurethane foam with a density of $\approx 0.05 \text{ g/cm}^3$ to simulate air cavities), internal air cavities, and the EBT-4 films (polyester, thickness 300 μm , density 1.35 g/cm^3) at the experimental measurement locations used for comparison with the film dosimetry data. Material properties were based on standard compositions or, for specific components like the foam, defined explicitly. For MARVIN simulations, AUTOMC performed the mapping of CT Hounsfield Units to material definitions (ABS, foam, air) and their respective densities. The inclusion of the film geometry was used to account for potential local perturbations to fluence and dose deposition at the film measurement planes...”

“Methods, Section 4.5

Quantitative dosimetry relied on EBT-4 radiochromic films (Ashland Inc., Bridgewater, NJ, USA), selected for their high spatial resolution and suitability for dosimetry in small, steep-gradient dose distributions...”

>> 9. The statement at Section 2 about the dose profile being narrower in the low-density insert irradiation is not clear. It might be better if there were a quantitative comparison instead of only a qualitative comparison based on the figure.

We agree that the original wording was too qualitative. To make the profile-width comparison clearer, we have added full width at half maximum (FWHM) values to the MC and film profile panels in Figs. 4 and 5. We have also revised the Results text to reference the key quantitative comparison directly. In the posterior posterior oblique plane seven-spot measurements, the measured right-left FWHM was slightly smaller for the low-density foam insert than for the ABS insert, 33.7 mm versus 36.2 mm, respectively, consistent with reduced lateral scatter in that direction.



Section 2.4.1

closely matched measured values, the higher foam peak reflecting reduced upstream energy deposition in the low-density insert. **FWHM values are reported on the profile panels to quantify the MC and film profile widths directly.** The 2-D gamma maps are uniformly green with only isolated red pixels at the periphery, with corresponding gamma pass rates of 98.67% and 99.94%

Section 2.4.2

...27.4 Gy, which was closely matched by the simulation (within $\pm 2\%$) (Fig. 5d). **FWHM values were added to the profile panels to quantify the width comparison directly. In the posterior oblique plane seven-spot measurements, the measured right-left profile was slightly narrower for the low-density foam insert than for the ABS insert (33.7 mm versus 36.2 mm, respectively; Fig. 5d,c),** consistent with reduced lateral scatter in that direction. The **posterior oblique-plane** 2-D gamma maps diverge:....

Reviewer #4:

>> I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Physics initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank the co-reviewer for their time, careful reading and constructive feedback.

Reviewer #1 (Remarks to the Author):

The manuscript has been modified accordingly.

We thank Reviewer #1 for reviewing the revised manuscript and for their positive assessment.

Reviewer #2 (Remarks to the Author):

After reviewing the revised manuscript, I find that the authors have adequately addressed all the concerns raised in the previous round of review. The manuscript has significantly improved and is now much clearer and easier to understand.

We thank Reviewer #2 for reviewing the revised manuscript. We appreciate their acknowledgement that the revisions have addressed the previous concerns and improved the manuscript's clarity.

Reviewer #3 (Remarks to the Author):

The manuscript has been considerably improved. This referee is satisfied with the responses to the previous round of comments, to the best of the authors ability, which have improved the scientific accuracy of the claims made in the paper.

We thank Reviewer #3 for reviewing the revised manuscript. We appreciate their positive assessment and acknowledgement that the revisions have improved the scientific accuracy of the manuscript.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Physics initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank Reviewer #4 for their contribution to the peer-review process.