Quality assurance phantoms for deep hyperthermia devices: design principles informed by computational modeling

Citation for the original published paper (version of record):
de Lazzari, M., Dobsicek Trefna, H., Carrapiço-Seabra, C. et al (2026). Quality assurance phantoms for deep hyperthermia devices: design principles informed by computational modeling. Physics in Medicine and Biology, 71(10).
http://dx.doi.org/10.1088/1361-6560/ae6748

Quality assurance phantoms for deep hyperthermia devices: design principles informed by computational modeling

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
30 June 2025

REVISED
28 March 2026

ACCEPTED FOR PUBLICATION
30 April 2026

PUBLISHED
15 May 2026

PAPER

Quality assurance phantoms for deep hyperthermia devices: design principles informed by computational modeling

Mattia De Lazzari[1,*], Hana Dobšíček Trefná[1], Carolina C Seabra[2], Patrick V Granton[2], Sergio Curto[2] and Dario B Rodrigues[3]

[1] Department of Electrical Engineering, Chalmers University of Technology, Göteborg, Sweden
[2] Department of Radiation Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
[3] Department of Radiation Oncology, University of Maryland School of Medicine, Baltimore, MD, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: lazzari@chalmers.se

Keywords: deep hyperthermia, experimental phantom, computational modelling, thermometry, SAR, treatment planning

Abstract

Objective. Accurate thermal dose delivery is essential for the clinical success of deep hyperthermia (DHT). As the European Society for Hyperthermic Oncology shifts toward temperature-based quality assurance (QA) metrics, standardized tissue-mimicking phantoms for DHT become increasingly important. This study provides quantitative evidence to guide phantom design using computational modeling supported by experimental validation. *Approach.* Numerical simulations were performed using a simplified model of a clinical DHT applicator operating at 75 and 100 MHz. Parametric studies assessed the influence of phantom geometry (diameter, length, wall thickness) and dielectric properties (relative permittivity and electrical conductivity) on specific absorption rate and temperature distributions. Agreement with patient models was evaluated by comparing phantom simulations with temperature profiles derived from anatomical models. A gel phantom was constructed to validate simulations experimentally. The effects of thermal mapping and positioning errors on thermal profiles were also analyzed. *Main results.* Simulations demonstrated that dielectric properties, especially electrical conductivity, had the strongest effect on heating patterns and temperature gradients; lower conductivity produced steeper focal profiles and reduced peripheral hotspots. Comparisons with patient models revealed good agreement in the focal region. Experimental measurements matched simulated temperatures near the applicator focus, with an average deviation of 0.7 ± 0.5 °C in the most reproducible series. Larger deviations near boundaries were attributed to thermal mapping uncertainties, catheter bending, and phantom misalignment. Including these uncertainties in the model indicated the need for margins reflecting average positioning errors of ± 2.5 cm along the probe axis and ± 0.5 cm perpendicular to it. *Significance.* This study provides a validated modeling framework and design recommendations that support temperature-based QA procedures for DHT. By clarifying how phantom parameters influence measurable temperature profiles and quantifying key sources of experimental uncertainty, this work strengthens the basis for standardized QA phantoms and enhances the reliability of performance assessment for clinical DHT systems.

1. Introduction

Hyperthermia therapy (HT) is an adjuvant cancer therapy that involves raising the tumor temperature to 40 °C–44 °C for approximately one hour. Numerous studies have demonstrated that HT acts as a potent biological sensitizer when combined with radiotherapy and/or chemotherapy (Horsman and Overgaard 2007, Oei *et al* 2020), leading to improved local tumor control and survival rates, with an excellent safety profile (Issels *et al* 2010, Lutgens *et al* 2010, Datta *et al* 2015, 2015, 2016, Van Driel *et al* 2018, Willner *et al* 2021). However, the clinical efficacy of HT is highly dependent on the accurate delivery of thermal

dose—that is, achieving and sustaining the required therapeutic temperature in the tumor for a prescribed duration, typically one hour (Leopold *et al* 1992, Franckena *et al* 2009, Bakker *et al* 2019, 2022a). This requires that HT systems generate spatially and temporally controlled heating while minimizing hotspots in surrounding healthy tissue. To meet these requirements, quality assurance (QA) guidelines have been developed over the years, focusing on the physical characterization of HT devices and their ability to produce controlled, reproducible heating patterns. In this context, the primary aim of this study is to support the design of standardized QA phantoms for deep hyperthermia (DHT) systems by providing the first systematic, quantitatively validated analysis of how phantom design parameters shape measurable temperature-based performance metrics.

Early QA approaches for DHT, such as those proposed by Lagendijk *et al* (1998), relied on visual inspection of luminous patterns generated by light-emitting diode (LED) matrix or lamp-based phantoms, as described in (Schneider and Van Dijk 1991, Schneider *et al* 1994). These phantoms typically consist of an array of LEDs or diodes submerged in a saline solution enclosed within a plastic shell—usually cylindrical or elliptical in shape. The saline solution is calibrated to replicate the electrical conductivity of human tissue, causing the LEDs or diodes to illuminate in response to the applied radiofrequency (RF) field. The resulting luminous pattern is then visually inspected to assess focus accuracy, steering, and field symmetry. Building on this approach, Bruggmoser *et al* (2011) introduced standardized QA guidelines recommending a cylindrical phantom with 20 cm diameter, filled with tissue-equivalent material consisting of a deionized water and salt mixture. The evaluation criterion in this method is based on energy deposition, quantified via the specific absorption rate (SAR) with units of W kg^{-1}. The applicator's performance was considered acceptable if the SAR ratio between the target and non-target volumes is at least 1.5.

More recently, the Technical Committee of the European Society for Hyperthermic Oncology (ESHO) has developed updated QA guidelines that emphasize temperature-based assessment of DHT systems (Dobšíček Trefná *et al* 2025). These guidelines represent a shift from earlier SAR-based metrics toward temperature-driven performance evaluation. Specifically, DHT systems are required to demonstrate a minimum temperature increase of 6 °C within 10 min in the focal region of a homogeneous tissue-mimicking phantom. In addition, the guidelines emphasize assessing the symmetry of the heating focus and the steering capabilities of the applicator, based on the temperature profiles measured within the phantom. To support this temperature-based QA approach, the guidelines recommend the use of standardized phantoms incorporating tissue-mimicking gels and embedded catheters to enable precise thermal mapping. The first experimental implementation of this updated methodology is presented in (De Lazzari *et al* 2024).

The present study contributes to this temperature-based QA approach by providing quantitative evidence and experimental validation for phantom design choices. The geometry of the QA phantom and the dielectric properties of the tissue-mimicking material are key determinants of RF energy deposition and the resulting temperature distributions. While the new 2026 ESHO DHT guidelines specify the use of tissue-mimicking phantoms for system verification, they do not analyze in detail how phantom design parameters influence the measurable heating profiles (Dobšíček Trefná *et al* 2025). Addressing this gap is important to ensure that QA measurements are reproducible across institutions and representative of typical applicator performance. In this study, we systematically investigate how variations in phantom dimensions and dielectric properties influence SAR and temperature distributions at clinically relevant frequencies (75 and 100 MHz). Phantom simulation results are compared with anatomical patient models and validated through controlled experimental measurements. To account for potential inaccuracies in temperature measurements, we repeated the experiments and subsequently examined numerically multiple sources of uncertainty expected to contribute to these, in particular thermal mapping errors, catheter misplacement, and phantom mispositioning. Altogether, this work provides evidence-based recommendations and validation results that complement the updated ESHO guidelines, clarifying how phantom design parameters affect heating patterns and strengthening the standardization of QA procedures for DHT devices. These efforts are consistent with recent roadmap publications highlighting the need for improved dosimetry, experimental validation, and standardized methodologies in non-ionizing radiation-based medical applications (Laakso *et al* 2026).

2. Materials and methods

All simulations in this study were performed using COMSOL Multiphysics v6.1 (COMSOL AB, Stockholm, Sweden), a commercial finite element numerical solver. The baseline computational model included a phantom positioned within the BSD-2000 Sigma 60 applicator (Pyrexar Medical, Salt Lake

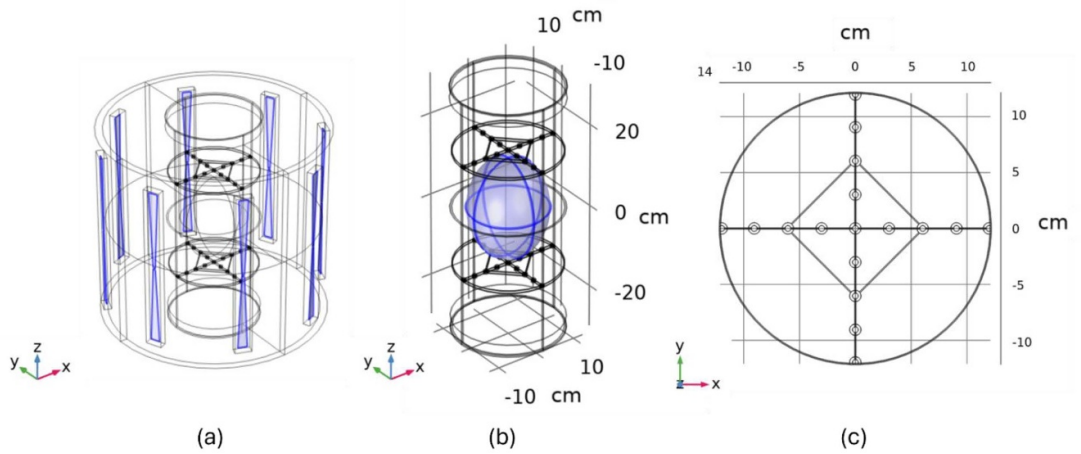

Figure 1. Computational model of the Sigma 60 applicator (Pyrexar Medical, USA) and ESHO QA phantom (a): 3D perspective view with antennas highlighted in blue surrounding the QA phantom; (b): 3D view of the QA phantom model (25 cm diameter), including the catheter holding structures and a central elliptical region ($12 \times 12 \times 18$ cm^3) used in mesh refinement; (c): close-up view of the catheter holding structure.

City UT, USA), a widely used clinical system for DHT. As described in the guidelines (Dobšíček Trefná *et al* 2025), the simulated standard QA experiment used a 10 min heating protocol with a total input power of 1000 W and a central focus targeting the geometric center of the applicator array (0, 0, 0) cm. Simulations were conducted at two clinically relevant frequencies (75 MHz and 100 MHz), where we evaluated the effects of phantom diameter, length, and wall thickness; dielectric properties of the tissue-mimicking material filling the phantom; and the influence of catheters used for thermometry and their supporting structures. To assess the relevance of these phantoms for QA, simulated temperature distributions were compared with those obtained in anatomical patient models. Finally, the numerical findings were validated through controlled experiments with a clinically operating Sigma 60 applicator.

2.1. Computational models

2.1.1. DHT applicator models

This study used a simplified model of the Sigma 60 applicator, illustrated in figure 1, which operates in the 75–120 MHz range. This applicator is configured as a phased-array system with eight equidistant antennas enclosed within a cylindrical, transparent plastic shell with an external diameter of 60 cm. A bolus membrane filled with deionized water is used as a coupling medium. The antennas are driven as four sets of adjacent pairs (top, bottom, left, right) and connected to four high-power RF amplifiers. The Sigma 60 antenna elements consist of bowtie-shaped planar dipoles with a total length of 43.5 cm and backed with a substrate (figure 2). The RF input signal is delivered to each antenna through an asymmetric coaxial feed, with a gap of 15 mm between the dipole arms. In our numerical models, the antenna geometry was simplified using a rectangular uniform lumped port instead of the coaxial feed, with the width corresponding to the average of both port edge widths, as shown in figure 2. The gap between the two dipole arms was also reduced from 15 mm to 5 mm. To reduce computational cost, the substrate and surrounding layers were removed, and the plastic applicator frame was replaced with water, so that the antenna was entirely immersed. While these simplifications affect the reflected power, they do not alter the normalized SAR pattern, which is the key parameter used in DHT treatment planning.

2.1.2. DHT phantom model

The QA phantom consists of a PVC tube equipped with catheters for housing temperature probes, which are secured by dedicated supporting structures (figure 1(c)). These structures are made of plastic and consist of annular hollow rings positioned at ± 14 cm from the phantom's central plane to minimize their impact on the SAR profile. The catheters are evenly spaced at 3 cm intervals along both the x- and y-axes (figure 1(c)). Once assembled, the phantom container is then filled with a tissue-mimicking gel prepared with a wallpaper paste powder, deionized water, and salt (Schneider *et al* 1995).

Two cylindrical phantoms with external diameters of 25 cm and 31.5 cm were modeled in COMSOL, with the 25 cm illustrated in figure 1. Parametric studies were conducted to determine optimal phantom length, wall thickness, and dielectric properties, as summarized in table 1. When sweeping the value

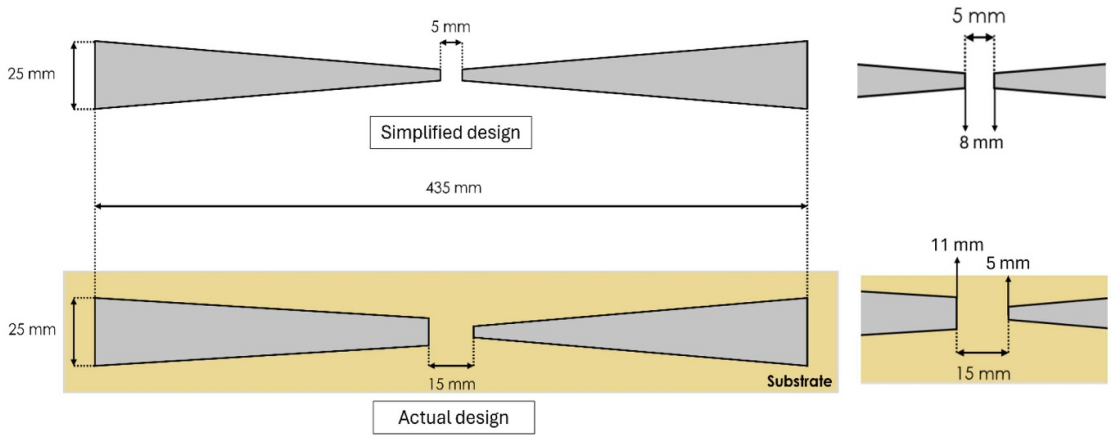

Figure 2. Antenna design used in the computational model (top) compared to the design implemented in the clinical Sigma 60 applicator (bottom). Right: close-up view of the antenna feeding port.

Table 1. Parameter ranges used in the phantom geometry and dielectric properties modeling studies, including their start and stop values and corresponding step sizes.

Parameter	Values/Range	Step	Nominal value
Phantom length (cm)	48, 54, 60–90	10	60
Phantom wall thickness (mm)	2–10	2	4
Gel relative permittivity (-)	40–70, 77.7	10	77.7
Gel electrical conductivity (S m^{-1})	0.3–0.6	0.1	0.6

of one parameter, the others were kept fixed at a nominal value. Simulations were performed at both 75 MHz and 100 MHz. To assess the influence of phantom length, we defined the metric SAR_{50}long, based on the SAR50 profile (i.e. the SAR 50th percentile of the SAR distribution), as the distance between the two SAR50 points along the z-axis. For the remaining design parameters listed in table 1, their influence on the temperature distribution was investigated along the x-axis on the central XY plane. The parameter values used for phantom fabrication were selected based on three criteria: minimizing perturbations to the phantom SAR field, maximizing observable temperature gradients, and maintaining a practical overall weight suitable for handling by two people.

The catheters used for temperature monitoring consist of tubes with a 4 mm outer diameter that cross the entire phantom length. Together with the catheter-holding structures, the catheters were incorporated in the computational model to evaluate their potential perturbation on SAR and temperature distributions, with three increasing levels of geometrical complexity: (1) solid PVC cylinders matching the 5.2 mm diameter of the catheter-holder holes; (2) solid PVC cylinders with the catheter's actual 2 mm outer diameter; and (3) a realistic catheter model consisting of a 0.25 mm PVC wall surrounding an internal air domain of 3 mm diameter. Although the DHT QA guidelines recommend evenly spaced catheter placement, we assessed their optimal positioning by determining the maximum observable temperature gradient between consecutive measurable locations along the x- and y-axes in the central transverse plane (figure 1(c)). This assessment was performed for both phantom diameters and frequencies.

2.1.3. Simulation settings

To mirror QA heating experiments, we performed coupled electromagnetic (EM)–thermal simulations to calculate the temperature distribution in the phantom during the 10 min heating experiment at 75 and 100 MHz. The computational model, shown in figure 1, consists of the Sigma 60 applicator with the phantom centrally positioned inside it. The simulation domain is truncated by an air-filled box surrounding the phantom and applicator, placed $\lambda_{air}/4$ from the antenna array ($\lambda_{air} = 3$ m, wavelength in air), where a scattering boundary condition (BC) was applied. The water bolus is supported by a silicone film, modeled as a transition BC with its actual 1.5 mm thickness. Dielectric properties are listed in table 2.

The heat transfer simulations were carried out only in the phantom domain, where we assigned convective heat flux BCs for both the water bolus/phantom and air/phantom interfaces. Heat transfer coefficients were set to 50 W m^{-2} K^{-1} and 5 W m^{-2} K^{-1}, respectively. The reference temperature in the computational modeling was set to 22 °C, corresponding to room and initial phantom temperatures.

Table 2. Dielectric and thermal properties used for the phantom and applicator models. For the tissue-mimicking gel, the modeled ranges are reported as the lower value, step size, and upper value.

Material	Relative permittivity ε (−)	Electrical conductivity σ (S m^{-1})	Thermal conductivity k (W m^{-1} K^{-1})	Heat capacity c (J kg^{-1} K^{-1})	Density ρ (kg m^{-3})
Air	1	0	0.0259	1005.5	1.293
PVC	3.2	$3.24 \cdot 10^{-3}$	0.16	950	1400
Water (22 °C)	78.4	$2.14 \cdot 10^{-3}$	0.60	4185	999
Gel phantom	40:10:70, 77.7	0.30:0.10:0.60	0.60	3900	1000
Silicone	3.2	0	0.20	1050	1100

Thermal properties are summarized in table 2. The 3D temperature distribution was calculated using the heat equation:

$$\rho c \frac{\partial T}{\partial t} = \nabla (k \nabla T) + \rho SAR$$

where ρ is mass density (kg m^{-3}), c (J kg^{-1} K^{-1}) is specific heat capacity, T (K) is temperature, t (s) is time, k (W m^{-1} K^{-1}) is thermal conductivity, and the external heat source induced by the RF field is given by ρSAR (W m^{-3}). The total input power of the applicator was equally divided among the four channels. Power was scaled to obtain a 6 °C temperature rise at the phantom center after 10 min heating.

2.1.4. Mesh independence study

Achieving mesh independence is essential to ensure the validity of a numerical solution. Thus, we determined the optimal mesh element size for each geometric domain within the model, including antennas, water, gel phantom, phantom walls, and catheter-holding structures. Our approach involved iteratively refining the mesh size for one domain at a time while keeping the mesh size constant for the remaining domains. This process was repeated until the solution converged, defined as a variation of less than 1% between refinement steps, as recommended in (Paulides *et al* 2021).

We defined distinct meshes for EM and thermal simulations by evaluating changes in SAR and temperature at the phantom center, i.e. coordinates (0,0,0) cm, respectively, which also corresponds to the central location of the target region used in all experiments. To ensure finer resolution in regions with higher expected spatial gradients, we added high-density mesh regions around the antenna elements and the target region. Thus, we added a 1 cm margin surrounding each antenna to capture the high RF electric field (E-field) gradients generated by the antennas and assigned a locally refined mesh to ensure accurate field resolution. Additionally, to account for the higher E-field and temperature gradients at the phantom center, we defined a target ellipsoid region ($12 \times 12 \times 18$ cm^3) in the middle of the phantom, as highlighted in figure 1(b). The mesh size was determined manually for thin elements, such as catheters, to preserve the geometric integrity of the computational domains.

The mesh independence study was implemented in COMSOL by assigning a mesh refinement parameter (m) to the maximum mesh element size and iteratively increasing it until convergence was reached. Table 3 summarizes the minimum and maximum element sizes assigned to each computational domain for both EM and thermal simulations. The table also presents the final optimized value of m at which solution convergence was achieved. As an example, figure 3 shows the convergence plots for SAR and temperature used to determine the optimal mesh size for the gel phantom domain, where we also included the total number of mesh elements (N_{mesh}), which is proportional to m. The catheters and their holders are omitted from table 3 because their mesh dimensions were dictated by their small size rather than their permittivity. Specifically, the minimum and maximum mesh element sizes for the catheter holders and solid catheters (no air) were set to 3 mm and 7 mm, respectively. When modeling the real catheter geometry (hollow cylinders), the maximum mesh element sizes was set at 1.5 mm and 2 mm for the catheter wall and internal air, respectively.

2.2. Comparison with anatomical patient models

To assess how well the proposed phantoms represent heating patterns expected in patients, we compared the temperature distributions obtained in the phantom with those from anatomical patient models. The aim was to evaluate how closely the phantom heating profiles resemble those observed in patients and to determine whether the phantom produces broadly comparable heating patterns. For this purpose, we selected CAD models of two cancer patients from the Erasmus Virtual Patient Repository (Bellizzi *et al* 2020)—Clarice and Will—each including a cervical or rectal tumor model, respectively. To reduce

Table 3. Minimum and maximum mesh element sizes set in COMSOL for each computational domain. The mesh refinement parameters (m) at convergence for thermal and EM simulations are also reported.

Computational domain	Mesh element size (mm)		Refinement parameter m at convergence	
	Min	Max	EM	Thermal
Dipole ports	—	$\frac{5}{2\cdot m}$	12.5	—
Dipole arms	—	$\frac{25}{2\cdot m}$	12.5	—
Water region surrounding dipoles	5	$\frac{\lambda\sqrt{\epsilon_{r,water}}}{2\cdot m}$	15	—
Remaining water region	—	$\frac{\lambda\sqrt{\epsilon_{r,water}}}{m}$	15	—
Phantom gel—central region	—	$\frac{\lambda\sqrt{\epsilon_{r,phantom}}}{m}$	17.5	7.5
Phantom gel—peripheral region	—	$\frac{\lambda\sqrt{\epsilon_{r,phantom}}}{2\cdot m}$	17.5	7.5
Phantom wall	1	$\frac{\lambda\sqrt{\epsilon_{r,phantom}}}{m}$	10	2.5
		$\frac{\lambda\sqrt{\epsilon_{r,phantom}}}{1.5\cdot m}$		

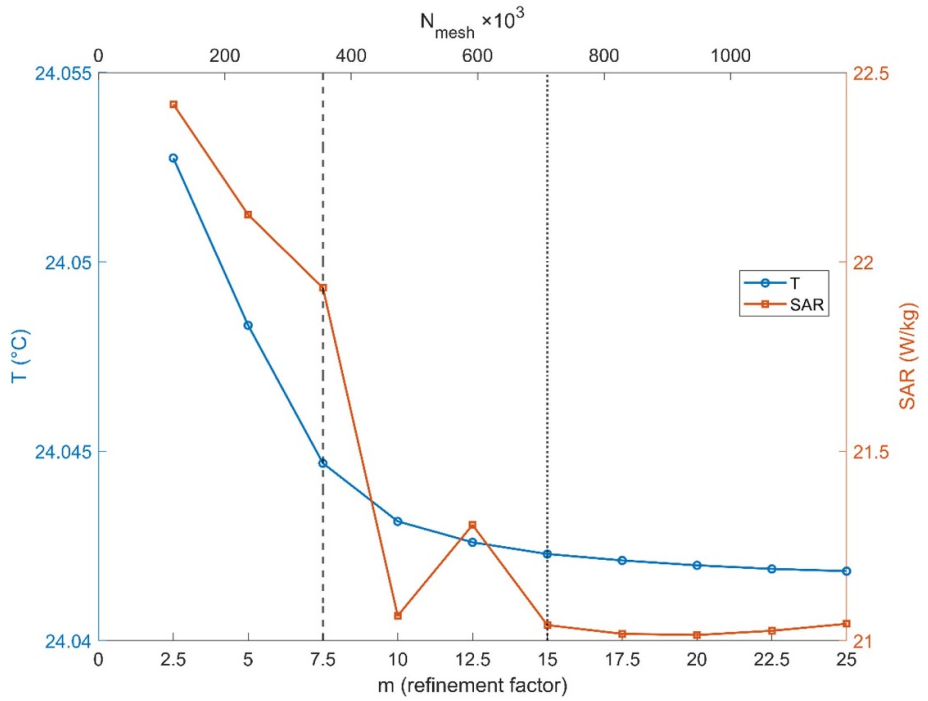

Figure 3. Convergence plots for SAR (orange) and temperature (blue) used to determine the optimal mesh size of the gel phantom. The SAR and temperature values at the phantom center decrease for increasing values of the mesh refinement parameter m. The total number of mesh elements (N_{mesh}) of the model is shown on the secondary x-axis. The selected value for m used throughout the simulations is marked with a vertical dashed (temperature) and dotted (SAR) lines.

the computational burden for the simulation of the interpolation grid, the patient models were downsampled to a 4 mm resolution using a winner-takes-all strategy (James and Sullivan 1992). A regular hexahedral mesh was then assembled, matching this resolution. The patient models were virtually positioned in the applicator by aligning the target region's central transverse plane with the center of the applicator (figure 4). The EM and thermal tissue properties for healthy tissues were retrieved from the IT'IS database (Hasgall *et al* 2022). Tumor properties were obtained as an average of all malignant-tissue properties reported by Joines *et al* (1994), as recommended by (Paulides *et al* 2021). These values are reported in table 4.

A SAR-based optimization was performed to determine the optimal power and phase of the four phased-array channels, with phase and amplitude of each channel used as optimization variables. We adopted the hot-to-cold spot quotient (HTQ) as the cost function, defined in (Zanoli and Dobšíček Tná 2022). The optimization procedure was implemented in MATLAB 2021 using an in-house developed HT treatment planning workflow (Zanoli *et al* 2023). The *E*-field distribution generated by each of the eight antennas was initially simulated in COMSOL. The convergence of the solution was verified by evaluating the antenna reflection coefficient (S11), ensuring that its value at the operating frequency remained stable with successive mesh refinements. The *E*-field distributions from each antenna were exported from

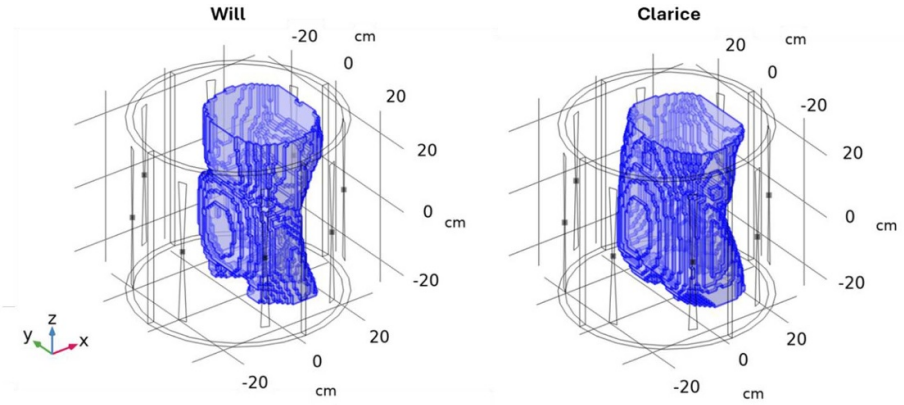

Figure 4. 3D view of the anatomical patient models Will (left) and Clarice (right) retrieved from the Erasmus Virtual Patient Repository (Bellizzi *et al* 2020) and positioned virtually within the applicator. The patient models are aligned such that the center of the target region coincides with the applicator center at coordinates (0,0,0) cm.

Table 4. Tissue dielectric properties (at 100 MHz), thermal properties, density and perfusion used for the patient models.

Tissue	Relative permittivity ε_r (−)	Electrical conductivity σ (S m^{-1})	Thermal conductivity k (W m^{-1} K^{-1})	Heat capacity c (J kg^{-1} K^{-1})	Density ρ (kg m^{-3})	Blood perfusion rate ω (ml min^{-1} kg^{-1})
Muscle	66.0	0.71	0.52	3421	1090	37
Fat	12.7	0.07	0.21	2348	911	33
Bone	15.3	0.06	0.28	2666	1908	10
Tumor	67.0	0.77	0.52	3421	1090	94.4

COMSOL to MATLAB and resampled to a uniform matrix with 4 mm spatial resolution and single precision. The SAR distribution was subsequently computed and further processed to enhance its correlation with temperature. This post-processing step included smoothing the SAR map using a 5 g mass-averaging scheme and excluding the voxels of the first 15 mm superficial tissues from the cost-function evaluation, i.e. these voxels were excluded from the HTQ calculation, since we can assume that they are cooled down by effect of the water bolus. The SAR-based optimization was then performed using the particle swarm optimization algorithm (Rijnen *et al* 2013).

The optimized SAR distribution was used as input for the subsequent thermal simulations, and the patient models were meshed similarly to the EM simulation. We then ran a transient simulation for 30 min, representing the time typically required to reach stable therapeutic temperatures in the clinic. The power was scaled to achieve a maximum temperature increase of 42 °C in healthy tissue using a local gradient descent optimization. The convection coefficient between the skin and water was set to 100 W m^{-2} K^{-1}, corresponding to the average value reported by van der Gaag *et al* (2006). The temperature distribution in the target volume (tumor) was characterized using the average and T50 and T90 indices, which represent the minimum temperature achieved within the highest 50% and 90% of the target volume, respectively.

2.3. Experimental verification

To validate the numerical studies, we performed QA experiments on the Sigma 60 applicator installed at the Erasmus MC Cancer Institute in Rotterdam, The Netherlands. We used a homogeneous tissue-mimicking phantom consisting of an external PVC cylinder with an outer diameter of 25 cm, wall thickness of 8 mm, and an overall length of 74 cm. A 3D representation of the phantom is shown in figure 5. The gel phantom was sealed with two waterproof lids at ±30 cm from the phantom center, yielding a total gel length of 60 cm. The lids consisted of a 2 cm-thick transparent acrylic disk with a 23.4 cm diameter (i.e. the internal diameter of the tube), tightly secured with plastic screws to a 2 cm-thick acrylic ring of the same diameter, which in turn was secured to the PVC cylinder with additional plastic screws.

The phantom was equipped with fifteen 6 French catheters (outer diameter ≈ 2 mm) to house temperature probes: 14 catheters oriented longitudinally along the vertical direction of the phantom (z-axis) and two radial catheters placed perpendicularly along the x- and y-axes. Each longitudinal catheter measured 33 cm in length, reaching the central plane of the phantom. Two longitudinal catheters

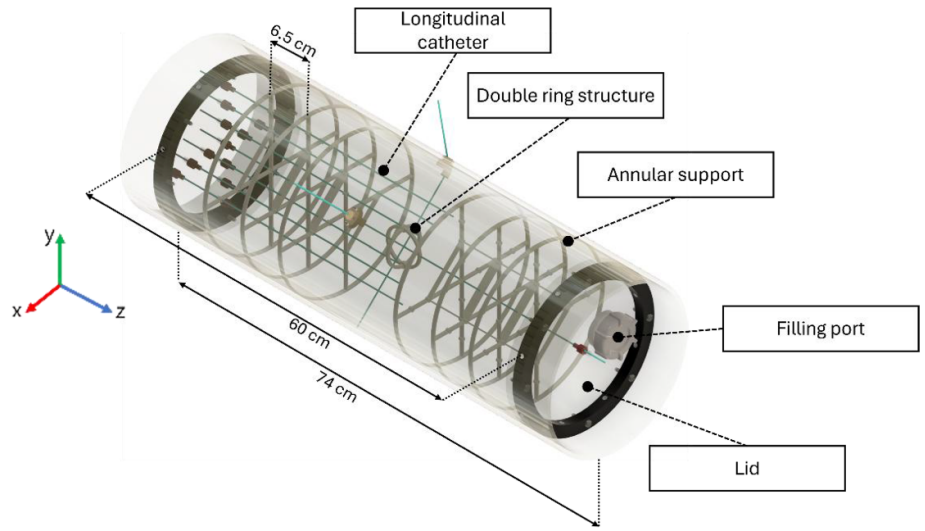

Figure 5. Schematic 3D representation of the phantom used during experiments, showing the external plastic shell, enclosing lids, catheters for temperature probes, and the 3D-printed holding structures. Adapted from De Lazzari *et al* (2024). CC BY 4.0.

crossing the entire phantom were installed along the z axis ($x = y = 0$ cm), accessible from opposite lids, enabling temperature measurements along the entire central z-axis. The catheter positions within the phantom were maintained using six equally spaced 3D-printed annular structures, positioned 10 cm apart, with the first located ± 10 cm from the phantom center. The two radial catheters intersected at the central axial plane of the phantom, forming a 90° angle with each other. To secure the two radial catheters, we used a smaller double 3D-printed ring structure with diameter of 6 cm, visible at the phantom center in figure 5. Another 3D-printed structure was incorporated at the phantom wall with a 2 cm radius curvature to guide the catheter into the axial plane without bending it. On the opposite end of the phantom wall, the closed tip of each radial catheter was securely attached to the PVC shell using water-resistant adhesive.

The phantom container was filled with approximately 30 l of a gel mimicking phantom, prepared by mixing deionized water with 3.5 g l^{-1} of salt, and 40 g l^{-1} of wallpaper paste powder (Metylan Normal, Henkel, Düsseldorf, Germany) at room temperature (21 °C). The solution was carefully mixed in 8 batches of 4 l each to minimize air entrapment. The solution was then poured into the container via the filling port and left to rest overnight with the port open to allow air to escape. The gap resulting from the mixture settling overnight was filled with the same gel mixture prepared in the day prior.

The dielectric gel properties (ε_r and σ) were characterized 24 h after preparation, once the gel mixture settled. One sample of about 70 g was taken from every other batch, resulting in four total samples used to verify batch consistency. Dielectric properties were measured using an open-ended coaxial probe (DAK 12, Schmid & Partner Engineering AG, Zurich, Switzerland) connected to a two-port VNA (ZNC 3, Rohde & Schwarz). Measurements were conducted at room temperature (21 °C) across the 50–500 MHz frequency range, with 1 MHz resolution. The probe was calibrated using an open-short-load routine, with deionized water at room temperature as the load. Calibration accuracy was confirmed by measuring a reference saline solution with known dielectric properties.

The gel thermal properties were measured in tandem with the dielectric properties using a commercial thermal analyzer (TEMPOS, Meter Group Inc., Pullman, WA, USA; accuracy: ±10%) with the dual-needle sensor model SH-3. These included the thermal conductivity (k), volumetric heat capacity (c) and thermal diffusivity (D). This meter uses the hot-wire method, where one needle applies heat for 30 s while a second needle records the resulting temperature rise for 90 s. Thermal properties are then computed by the device as described by Silva *et al* (2020), Farina *et al* (2020). The density of the phantom was calculated from these thermal properties according to the equation $\rho = k/(cD)$.

Before experiments, a computed tomography (CT) scan of the filled phantom was acquired to verify the correct positioning of the catheters and to exclude the presence of large (>1 cm) air bubbles. The phantoms were then positioned in the applicator using custom-made wooden supports to ensure the correct vertical position of the phantom relative to the applicator aperture. A laser positioning system was used to assess accurate alignment of the phantom along the applicator longitudinal axis (z-axis).

The temperature was monitored using six single-sensor thermistor probes and two multi-sensor fiber optic temperature probes available at the clinical site. Four of the thermistors were used for thermal mapping (i.e. automated mechanical movement of the probe within the catheter), while the remaining two were fixed at stationary positions. The fiber optic probes were equipped with six sensors with 2 cm spacing, thus covering a length of 10 cm. Before proceeding with the experiment, the probes were calibrated using a temperature-controlled water bath to ensure a maximum error of ± 0.2 °C relative to a calibrated reference. Additionally, the accuracy of the thermal mapping system was verified through a dry run scan in a marked transparent catheter (1 cm resolution), measuring the deviation between the planned mapping position and the measured one. Two mapping probes were inserted into the radial catheters, and the other two were inserted into the central longitudinal catheters, from the two opposite ends of the phantom. The two remaining static thermistors were taped to the external surface of the phantom—at the top and bottom, aligned with the mid-axial plane—to measure temperature at the bolus-phantom interface. The fiber optic probes were inserted into the longitudinal catheters at coordinates $(x, y) = (6, 0)$ and $(0, 6)$ cm. To ensure that the tips of the stationary probes were positioned at the central plane of the phantom, we measured the respective intra-catheter lengths and placed tape markers at the catheter entry points.

We planned two series of three identical experiments, but the first series had to be reduced to two experiments due to a conflict with the clinical schedule. We followed the same measurement procedure in all repetitions: an initial baseline thermal mapping scan was performed with the power off, using a 1 cm step size, a 20 cm mapping length, and a dwell time of 6 s, which took 2 min to complete. Then, the power was turned on for 10 min with a total forward power of 1000 W, and zero phase settings to obtain a central focus at (0,0,0) cm. Immediately after power was turned off, a second thermal mapping scan was executed to record the resulting post-heating temperature profile. Following each measurement, the phantom was allowed to equilibrate with room temperature for at least 8 h before reuse.

The experimental validation in this study was limited to the central-focus configuration of the applicator (0,0,0 cm). This configuration represents the standard reference setting used in QA procedures and minimizes phase steering effects that could introduce additional uncertainties in the interpretation of phantom measurements. While off-center steering conditions are clinically relevant, their validation would require substantially larger experimental campaigns and was therefore beyond the scope of the present study.

The equivalent computational model of the aforementioned heating experiment was simulated using the same numerical methods described in section 2.1, where the geometry, dielectric, and thermal properties of the phantom model were adjusted to match experiments. Additionally, the double annular structure of the catheter-holding structure and the two 3D-printed inlets for the radial catheters were incorporated into the model to verify their impact on SAR and temperature distributions. The simulated power was calibrated to match the maximum experimental temperature increase at the phantom center for each measurement. The individual channel phases used in the simulations were set as the average phase recorded by the DHT control software. During experiments, we observed that the thermal mapping position had significant uncertainties, which we accounted for by simulating the maximum and minimum temperature increase within a cylindrical volume (1 cm in diameter and 5 cm in length) centered at each experimental mapping location. Additionally, we assessed the impact of phantom misalignment relative to the applicator by simulating ± 1 cm shifts along the x-, y-, and z- axes.

3. Results

3.1. DHT phantom model selection

Table 5 presents the simulated maximum SAR and temperature in the phantom target region for four different simulation scenarios with increasing complexity in the modeling of catheter structures. Scenario A excludes the catheters and includes only the catheter holders, Scenario B adds the catheters as a solid domain with a radius of 2.6 mm (i.e. completely filling the corresponding gaps in the annular support structure), Scenario C uses realistic catheter dimensions with a 2 mm radius, and Scenario D provides the most accurate representation of the actual catheter structure, as it includes the catheter wall and inner air domain. The table also includes the corresponding number of mesh elements and simulation time. We computed the percentage differences in SAR and temperature maximum values between the accurate Scenario D and the simplified Scenarios A-C. The results indicate a maximum deviation of 3% compared to Scenario D, confirming that all considered models provide a close approximation of the accurate scenario. Based on these results, subsequent simulations were performed using Scenario A, which includes only the holder structures and provides the fastest computational time.

Table 5. Simulation time, mesh element count, and maximum SAR and temperature in the phantom's elliptical central region for four simulated scenarios with different catheter modeling complexity. Percentage differences are shown for the first three scenarios compared to the accurate reference geometry in Scenario D.

Scenario	Description	Simulation time (h:min)	# Mesh elements ($\times 10^6$)	Max SAR (W kg^{-1})	ΔSAR (%)*	Max T (°C)	ΔT (%)*
A	Just holders	1:33	1.14	40.9	1%	28.1	0%
B	2.6 mm catheters, full solid	7:02	1.91	41.6	3%	27.9	0%
C	2 mm catheters, full solid	5:53	1.77	40.4	0%	27.9	0%
D	2 mm catheters, with air gap	24:33	3.75	40.3	—	28.0	—

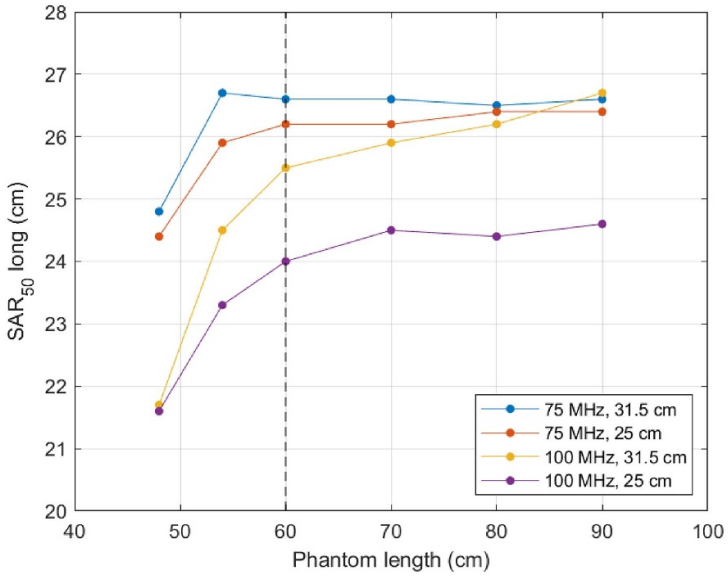

Figure 6. Effect of phantom length on the SAR_{50}long parameter at 75 and 100 MHz for phantom diameters of 25 and 31.5 cm. The dashed line indicates the phantom length selected for the QA experiments.

3.2. Phantom parametric studies

Figure 6 depicts how SAR_{50}long varies with different gel phantom lengths across both diameters and frequencies. Since increasing phantom length would significantly increase its weight, we selected a phantom length of 60 cm (as shown by the dashed line in figure 6). This length also allows the standard BSD2000 temperature probes to reach the phantom center.

The effects of wall thickness, relative permittivity, and electrical conductivity were analyzed by examining the 1D radial temperature profile along the x-axis of the central transverse plane. Figure 7 shows the temperature distribution for different wall thicknesses, figure 8 illustrates the influence of permittivity, and figure 9 depicts the impact of electrical conductivity. Each analysis is reported for both 75 MHz and 100 MHz. In each case, the nominal value is indicated by a dashed line. Variations in wall thickness have a negligible impact on temperature gradients at the phantom center but show a small effect near the phantom wall. In contrast, lower relative permittivity values degrade temperature gradients, intensifying peripheral heating near the phantom wall, particularly in the 31.5 cm phantom. Conversely, decreasing conductivity improves temperature gradients, which can facilitate temperature measurements and characterization of the focal region due to the higher temperature difference between measurement points. With lower conductivity, superficial hot spots are also reduced.

The optimal placement of catheters for temperature probes was determined by examining the radial 1D temperature profile in the central axial plane of the phantom. The maximum temperature gradient observed between consecutive probes was 0.5 °C at 75 MHz and 0.9 °C at 100 MHz. Figure 10 presents the temperature profiles for both diameters and frequencies, with the optimal catheter locations marked by a dashed line for 75 MHz and a dotted line for 100 MHz.

Figure 11 compares the simulated 1D radial temperature profiles in the central axial plane of the phantom for different electrical conductivity values with the corresponding 1D temperature profile from

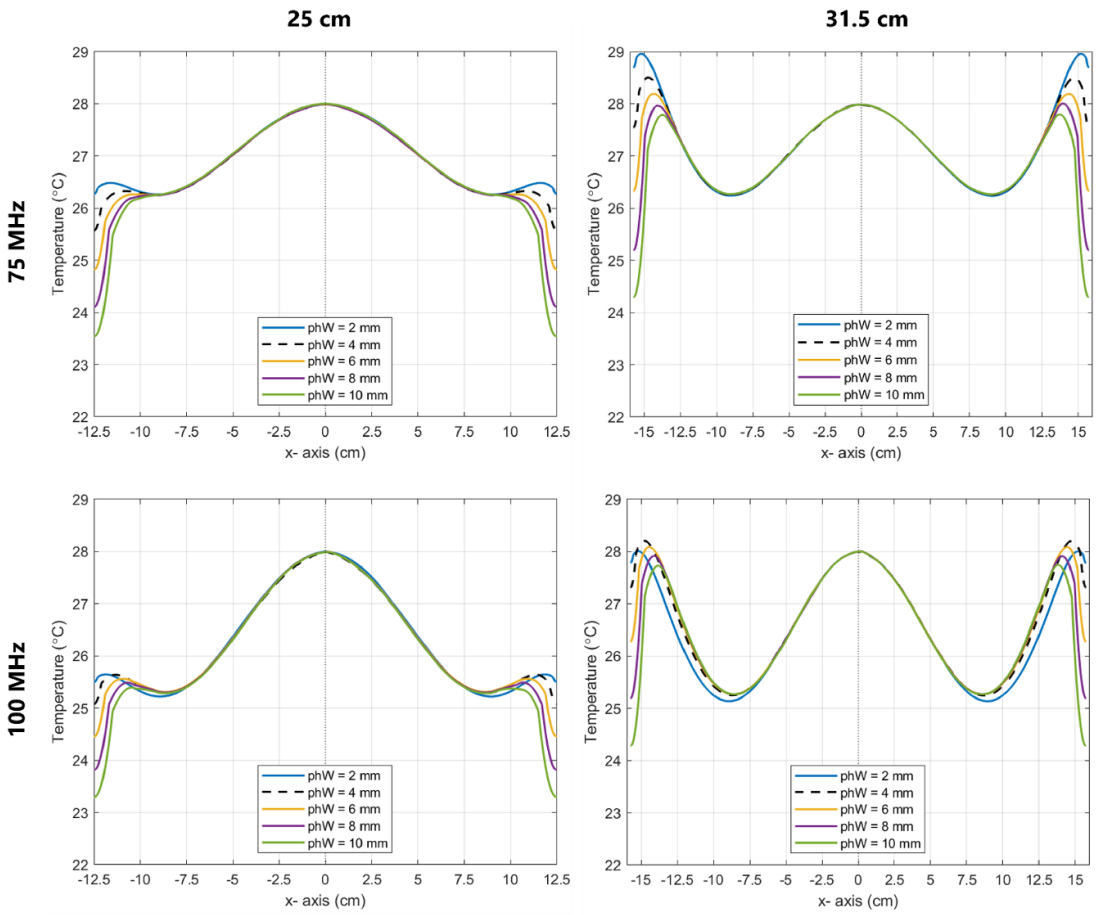

Figure 7. Impact of the phantom wall thickness (phW) on the temperature distribution along the x-axis at the phantom central axial plane, for different frequencies and phantom sizes. The dashed line indicates the nominal value of phW = 4 mm.

patient models simulations along the *x*-axis passing through the target area. Figure 11 also shows a comparison of longitudinal temperature profiles along the vertical *z*-axis for both patient and phantom models. The results indicate good agreement at the center, while deviations between the two are observed toward the edges, with differences within 20%.

3.3. Experimental verification

3.3.1. Phantom properties characterization

The measured density, dielectric, and thermal properties of the gel phantom are presented in table 6. For reference, the table also includes the human muscle properties retrieved from the IT'IS tissue properties database (Hasgall *et al* 2022) together with the percentage differences between the measured phantom values and the IT'IS values. It should be noted that the phantom is intended to represent the average dielectric properties of human tissues in the pelvic region; therefore, an exact match to any single tissue type is neither expected nor required. In clinical practice, most clinical centers operate at frequencies between 90 and 105 MHz. Within this range, the variation in dielectric properties is limited to 0.1% and 0.3% and can be considered negligible.

3.3.2. Phantom temperature profiles: comparison of measured and simulated data

We completed five repetitions of the same QA heating experiment. However, due to clinical scheduling constraints, only the last three repetitions were performed without replacing the phantom in the applicator, thereby ensuring the exact same experimental setup. The experiments were performed at room temperature (20.2 $\pm$ 0.3 °C) with an initial water bolus temperature of 21.5 $\pm$ 0.3 °C. The average phase recorded for the applicator four channels was consistent across all repetitions, with a maximum phase variation of 1° across the five measurements.

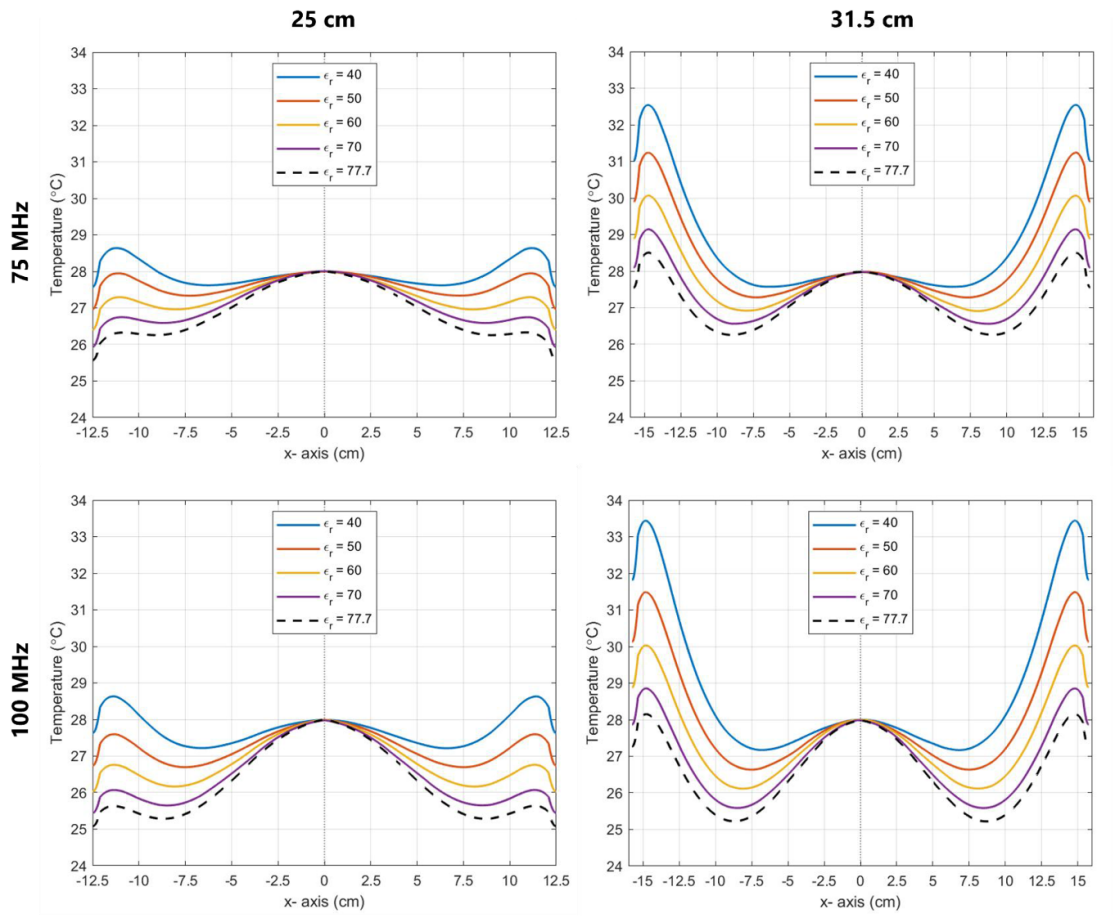

Figure 8. Impact of the phantom gel relative permittivity (ε_r) on the temperature distribution along the x-axis at the phantom central axial plane, for different frequencies and phantom sizes. The dashed line indicates the nominal value of $\varepsilon_r = 77.7$.

Figure 12 shows the temperature increase (ΔT) profiles measured using the thermal mapping probes, both longitudinally and axially, compared with the corresponding simulation results, including uncertainties due to thermal mapping (top row—a and b) and phantom positioning (bottom row—c and d). The measured temperature profiles were corrected using the calibration procedure discussed in section 2.3. The dotted lines represent experimental data from the first two measurements (Series S1), the dashed lines from last three measurements (Series S2), and the solid lines denote the simulated profile. For reference, temperatures measured by the fiber-optic probes are shown as two points located at $\pm$6 cm together with their measurement variability.

The average difference between the measured and simulated ΔT in figure 12 was 0.9 $\pm$ 0.5 °C for S1 and 0.7 $\pm$ 0.5 °C for S2. However, closer inspection of the longitudinal temperature profiles reveals that the maximum temperature is not located at the phantom center, but approximately 2 cm off-center. This artificial temperature plateau is very likely caused by probe drag after the probe became stuck during mapping. As a result, the applied power required to achieve the maximum temperature increase along the longitudinal axis is clearly overestimated, which manifests as an overprediction of temperature in the axial plane measured by the axial probes. When the applied power is reduced accordingly, the agreement between measured and simulated temperatures in the axial plane improves substantially, particularly for series S2.

To evaluate the impact of thermal mapping inaccuracies and phantom positioning, we carried out additional simulations by displacing the phantom and/or temperature probes from their nominal positions. The blue band in the top row represents the uncertainty associated with thermal probe positioning ($\pm$2.5 cm along the probe axis and $\pm$0.5 cm perpendicular to the probe axis) and corresponds to the range between the minimum and maximum simulated temperatures within the evaluated cylindrical

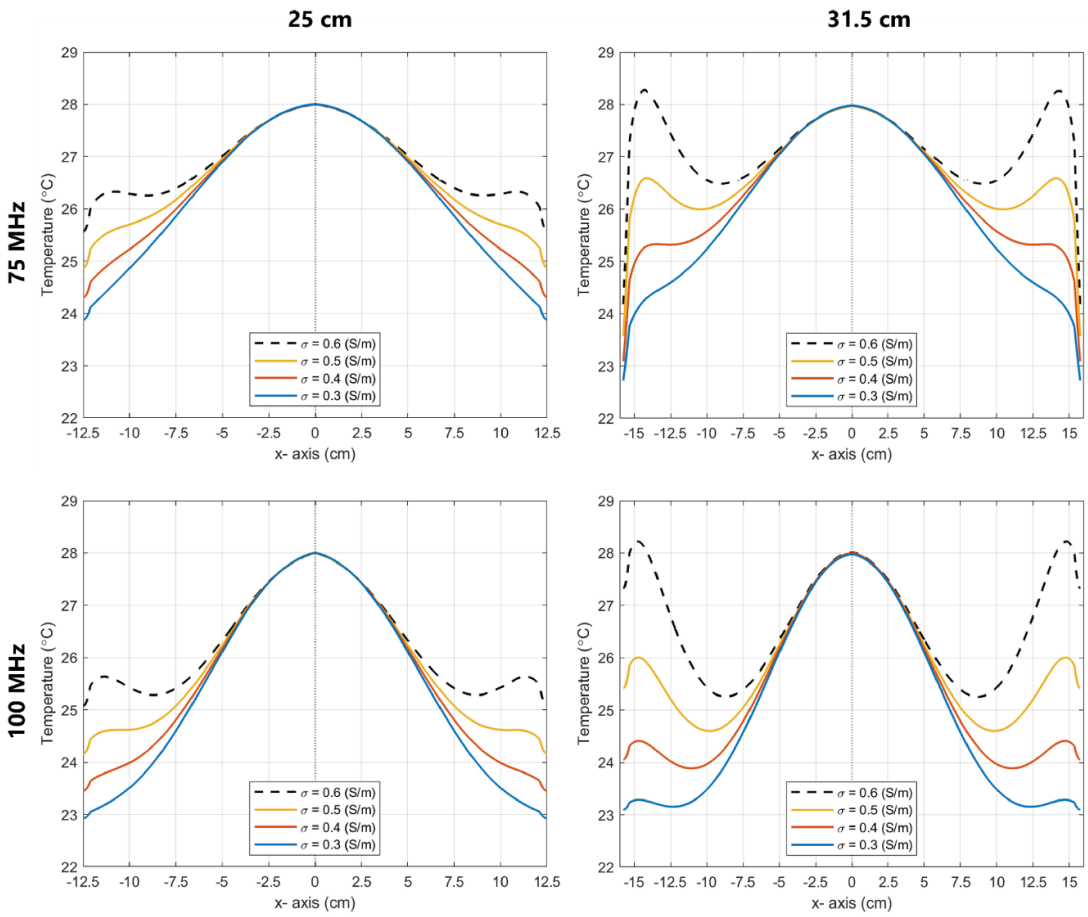

Figure 9. Impact of the phantom gel electrical conductivity (σ) on the temperature distribution along the x-axis at the phantom central axial plane, for different frequencies and phantom sizes. The dashed line indicates the nominal value of $\sigma = 0.6$ S m^{-1}.

volume. The violet color reflects the combined effect of uncertainties in both power variation and probe positioning. The yellow band in the bottom row figure 12 represents the uncertainty associated with phantom misalignment and reflects the range of simulated temperatures at the measurement locations for phantom displacements of ±1 cm in all directions.

4. Discussion

The findings of this study clarify how phantom geometry, dielectric properties, and catheter-support structures influence the temperature patterns generated by a clinical phased-array DHT system. By integrating systematic parametric simulations with experimental verification, we quantified how these design choices affect focal and peripheral heating as well as the reproducibility of measurable thermal gradients. While the updated ESHO QA guidelines present the final recommended deep-hyperthermia QA phantom, the foundational modeling, parameter optimization, and experimental validation that underpin that design are presented exclusively in this work. We also compared phantom-based temperature profiles with those obtained from anatomical patient models to assess clinical relevance. The following discussion interprets these results in the context of QA phantom design and their practical implementation under the updated ESHO guidelines for DHT systems.

4.1. Rationale for phantom design selection

The design of the QA phantom needed to satisfy both functional and practical requirements: it had to enable accurate assessment of the temperature distributions generated by the applicator while also remaining easy to fabricate, handle, and implement in routine QA workflows. The phantom diameter was chosen based on the operating frequencies of common DHT applicators. A larger diameter of 31.5 cm was selected for use with lower-frequency systems, particularly those operating around 70 MHz, whereas a smaller diameter of 25 cm was selected for higher-frequency systems (80–120 MHz). This size

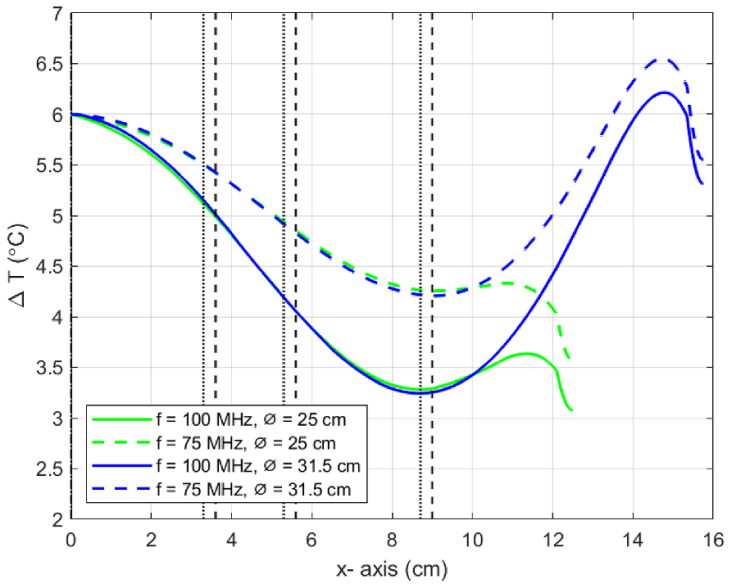

Figure 10. Simulated temperature rise along the positive x-axis at the central axial plane of the phantom after 10 min of heating at 75 MHz and 100 MHz for two phantom diameters (∅). Dotted vertical lines indicate the optimal catheter positions for ∅ = 25 cm, and dashed lines mark those for ∅ = 31.5 cm.

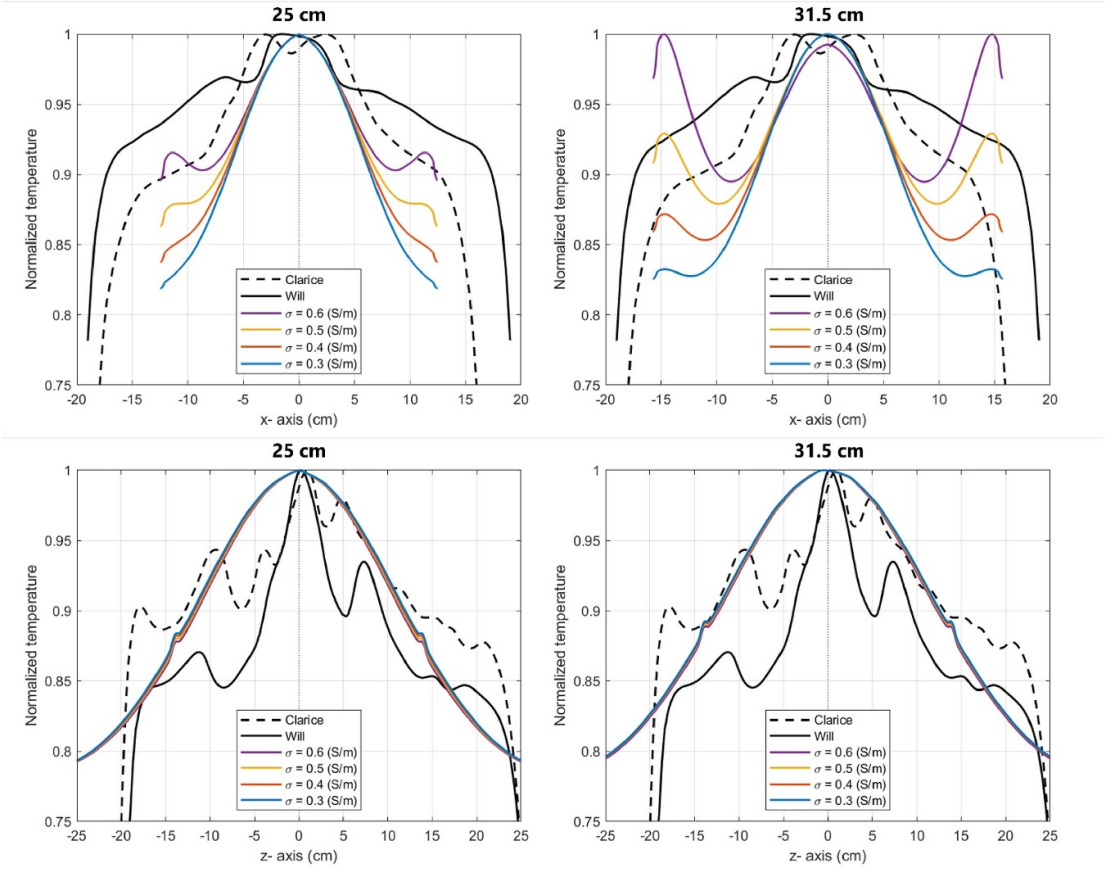

Figure 11. Normalized temperature distribution along the x-axis (top row) and z-axis (bottom row) for phantoms with varying conductivity values, and for the two patient models: Clarice (dashed black line) and Will (solid black line).

is compatible with standard PVC tubing readily available worldwide and closely corresponds to common tubing diameters used in regions with the imperial system (10 in = 25.4 cm and 12.5 in = 31.8 cm), facilitating material sourcing and fabrication. Although standard PVC wall thicknesses range from 2 to

Table 6. Measured properties of the muscle-mimicking phantom at 100 MHz and associated uncertainties. For comparison, the density, dielectric, and thermal properties of human muscle tissue from the IT'IS database are also listed. The percentage differences between the measured values and the IT'IS reference data are shown in the last row.

		Dielectric properties		Thermal properties	
	Density (kg m^{-3})	Relative permittivity ε_r (−)	Electrical conductivity σ (S m^{-1})	Specific heat capacity c (J kg^{-1} K^{-1})	Thermal conductivity k (W m^{-1} K^{-1})
Phantom (measured)	1001 ± 2.1	72.7 ± 1.1	0.61 ± 0.02	3693 ± 83	0.60 ± 0.01
Muscle (Hasgall *et al* 2022)	1090 ± 55	66 ± 3.3	0.708 ± 0.04	3421 ± 460	0.49 ± 0.04
Difference (%)	8	10	13	8	20

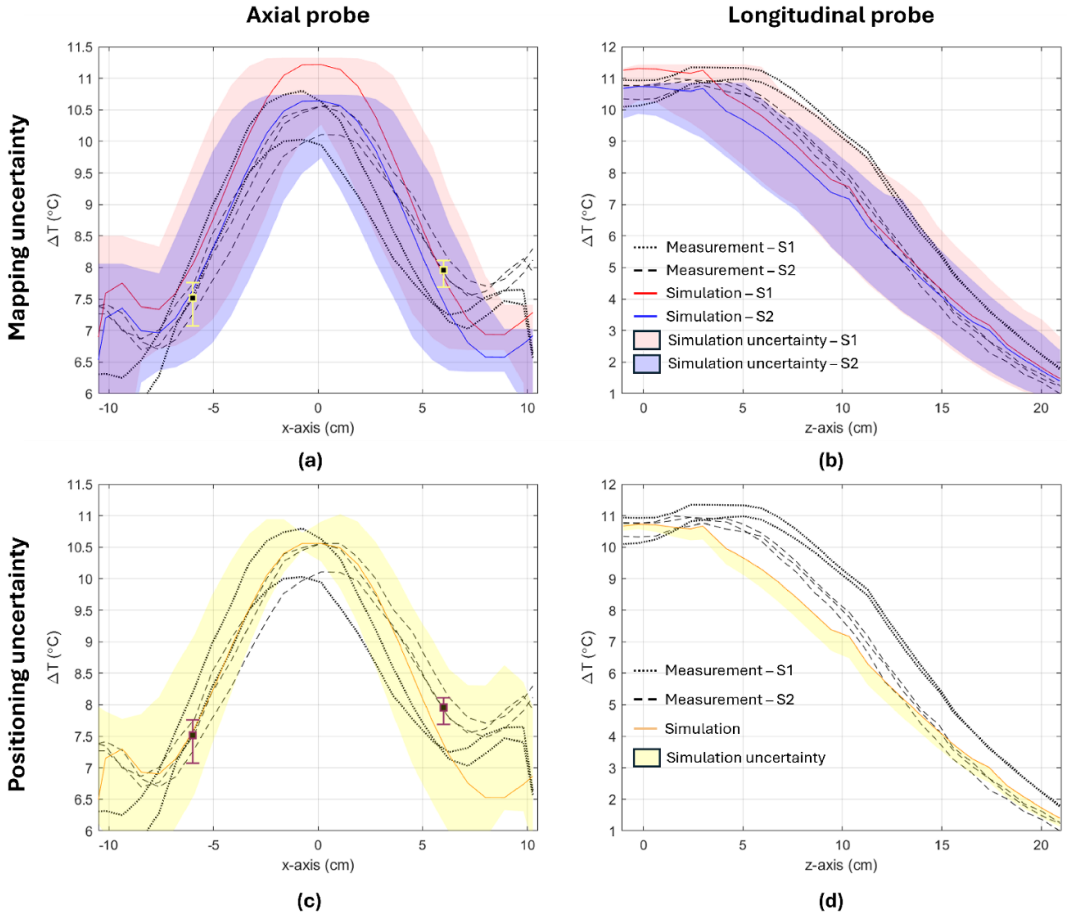

Figure 12. Simulated vs. measured temperature distribution in the axial (a), (c) and longitudinal phantom directions (b), (d), including experimental series 1 (S1) and 2 (S2). Top row: The red and blue shaded areas represent the range between the minimum and maximum temperatures simulated within a cylindrical volume centered on each measurement point to account for thermal mapping position inaccuracies for series S1 and S2, respectively. Bottom row: The yellow shaded area represents the range of temperatures simulated at the measurement points across all considered phantom misplacement of ±1 cm, reflecting the effect of phantom-positioning errors.

15 mm, our simulations demonstrated that such variations have a negligible impact on the central temperature gradients.

The phantom length was initially set at a minimum of 48 cm, which corresponds to the Sigma 60 applicator length. However, to mitigate standing wave effects and avoid interference from boundary reflections, we evaluated the longitudinal SAR distribution along the phantom length. As shown in figure 6, the SAR50 profile stabilized at both tested frequencies and diameters for phantom lengths above 60 cm, indicating minimal influence from wave reflections beyond this length. Notably, the SAR50 profile did not stabilize with increasing phantom length for the 31.5 cm phantom at 100 MHz, but this larger phantom is primarily intended for use at lower-frequency applicators (70 MHz). These findings also informed the placement of catheter support structures: to avoid disturbing the EM field in the central region, supports should be positioned at least 10 cm from the phantom center.

The phantom container was filled with gel mixture based on a wallpaper paste, which is one of the most commonly used tissue-mimicking materials for hyperthermia QA. First introduced in 1995 (Schneider *et al* 1995), its properties—particularly viscosity and electrical conductivity—can be tuned by adjusting the concentrations of wallpaper paste powder and salt, respectively. Its dielectric and thermal characteristics have been well documented (Schneider *et al* 1995, Farina *et al* 2020), making it a reliable and reproducible choice for hyperthermia QA. In our work, this formulation not only produced dielectric values within the expected range but also yielded stable and spatially uniform thermal behavior, which contributed directly to the close agreement observed between simulations and measurements in the focal region. This practical robustness underscores why this gel mixture remains well suited for standardized, reproducible QA procedures across DHT centers.

The standard gel formulation aims to replicate the dielectric properties of muscle tissue, with a nominal conductivity of approximately 0.66 S m^{-1} and a relative permittivity of 77.7 at 100 MHz. However, our simulation results showed that reducing the conductivity enhances the steepness of thermal gradients—a desirable feature for evaluating temperature-based performance metrics. This effect was particularly pronounced in the 31.5 cm phantom, where peripheral hotspots became more prominent at both tested frequencies. For this reason, using a reduced-conductivity gel mixture may be advantageous. For instance, a conductivity of 0.5 S m^{-1} was sufficient to suppress peripheral temperature peaks while maintaining the maximum temperature increase observed at the phantom center (figure 9).

While muscle tissue is often used as a reference for phantom fabrication, it is worth noting that most DHT applicators in clinical use target abdominal and pelvic tumors—regions characterized by a heterogeneous mixture of tissues, including fat, bone, and gas-filled bowel, which have significantly lower electrical conductivity. Therefore, designing the phantom to mimic the volume-averaged dielectric properties of this tissue composition may improve clinical relevance. As an approximation, using two-thirds of muscle's conductivity yields a target value of 0.44 S m^{-1} at 100 MHz.

Another key phantom design consideration is the placement of the catheters that guide the temperature probes inside the phantom. Their position was optimized to maximize the spatial resolution of observable temperature gradients within the phantom. Based on this criterion, we identified the optimal locations for longitudinal probes, as shown in figure 10. In the experimental phantom—originally developed for a separate study—we used a regular grid with 3 cm spacing. This configuration may be easier to manufacture, but we recommend adopting the optimized layout derived from our analysis, with catheters placed at $\pm$3.4 cm, $\pm$5.3 cm, and $\pm$8.7 cm along the *x*- and *y*- axes for the 25 cm phantom; and at $\pm$3.7 cm, $\pm$5.7 cm and $\pm$9 cm for the 31.5 cm phantom. Since the catheter holder is a 3D-printed component, it can be readily modified to accommodate alternative probe layouts.

4.2. Phantom vs. patient temperature simulations

The primary goal of the proposed phantom was not to replicate patient anatomy or reproduce the exact temperature distributions observed in clinical treatments. For such purposes, anatomically realistic anthropomorphic phantoms have been developed and described elsewhere (Curto *et al* 2019). The phantom was also not designed to address patient-safety aspects, which should be assessed through patient-specific treatment planning and comprehensive thermometry during treatment delivery. Instead, this phantom is intended to serve as a standardized, reproducible tool for evaluating the thermal performance of DHT systems in a controlled and setting. Nevertheless, we still aimed to obtain clinically relevant temperature increase distributions in the proposed QA phantom. As shown in figure 11, the relative temperature distributions in the phantom and patient models are well-aligned at the applicator's focal region. Toward the periphery, this agreement decreases, which is expected due to the phantom's simplified cylindrical geometry and homogeneous medium, which contrast with the contoured and heterogeneous patient anatomy. Despite this, the relative temperature profiles are qualitatively comparable and fall within the upper 20% range of the normalized temperature, supporting the phantom as an adequate experimental surrogate patient model.

Interestingly, the temperature distribution in the 25 cm phantom filled with a gel conductivity of 0.6 S m^{-1} (representative of muscle tissue) closely matches that observed in the patient models. This finding contrasts with the earlier discussion in section 4.1, suggesting that reducing the phantom's conductivity can improve temperature gradient steepness and better approximate the average dielectric properties of abdominal and pelvic tissues. This discrepancy underscores the importance of clearly defining the phantom's intended purpose. Specifically, our goal is to optimize the phantom for QA purposes, a lower conductivity may be preferred to increase thermal gradient resolution and enhance performance benchmarking. In contrast, if the objective is to approximate clinically observable heating patterns, a higher conductivity, consistent with muscle tissue, may offer a closer match to patient data—albeit with

reduced sensitivity to spatial gradients. It should also be emphasized that the phantom can only realistically reproduce the relative temperature distribution observed in patients. An absolute temperature match cannot be expected, as thermal conduction in the gel alters the gradients that would otherwise be observed. This distinction highlights that phantom design should be guided by its intended use—whether for routine QA testing or for approximating clinically observable heating patterns in preclinical system evaluation. Importantly, the use of this phantom is limited to the characterization of system performance in terms of heating capability. It is not intended to verify clinical effectiveness or patient safety. Those aspects must instead be evaluated using detailed anatomical phantoms, patient-specific treatment planning, and clinical thermometry. The present phantom is designed to verify that the system can deliver the intended amount of heating and accurately focus and steer the energy deposition toward the target region. When this capability is confirmed and combined with appropriate patient-specific treatment planning, reliable and effective clinical treatment delivery can be achieved.

4.3. Experimental verification of numerical modelling

While guidelines for performing hyperthermia computational modeling are available (Schneider *et al* 1994), formal verification procedures for assessing model accuracy remain limited. Experimental data obtained using tailored, rigorously designed phantoms can serve this purpose. In this study, we compared the simulated temperature rise with that measured in a physical phantom using a well-established clinical DHT device. To account for potential inaccuracies in temperature measurements, we repeated the experiments and subsequently numerically examined multiple sources of uncertainty expected to contribute to these, in particular thermal-mapping position errors as well as catheter and phantom positional deviations.

Thermal mapping errors emerged as the dominant source of uncertainty. Although we performed a calibration for the mapping probe position, this was done using a setup where the probe cables were not under any tension. In the actual experimental setup, with the probe cables routed to allow thermometry studies at different locations, we realized that different probe cables had different tensions (e.g. from being stretched to reach the other side of the phantom). This led to a noticeable inconsistent probe drag during mapping, particularly during the first 10 cm of mapping. We attempted to quantify the resulting positional offset by placing a ruler in the thermal mapping unit adjacent to the moving cable and found that the probe position deviated by several centimeters (estimated to be at least 3 cm) from their planned locations. An additional uncertainty was introduced by catheter deformation: phantom CT scans revealed that internal pressure from the tissue-mimicking gel caused some catheters to bend by up to 0.5 cm from their intended paths. To incorporate both sources of uncertainty into the simulations, we defined a sub-volume centered on each nominal measurement location, with dimensions reflecting average positional uncertainties: $\pm$1.5 cm along the probe axis (to capture thermal mapping uncertainty) and $\pm$0.5 cm perpendicular to the axis (to capture catheter bending).

Accounting for positional uncertainties improved agreement between simulated and measured temperatures, particularly near the boundaries. However, this strategy was less effective for the longitudinal temperature profiles, likely because probe drag during mapping along that direction was more pronounced and unpredictable than assumed in the model. In the first measurement series (S1), the longitudinal profiles clearly show evidence of pronounced probe drag: the probe appears to have become stuck during mapping, producing an artificial temperature plateau. Indeed, shifting the simulated curve by approximately 2 cm to the right results in a substantially improved match. The unreliable mapping is also evident in the axial profiles, where the curves show a systematic shift to the left. In this case, probe calibration offers limited benefit because the drag occurred during the measurement acquisition, not during calibration. Therefore, for the first measurement series (S1), the assumed deviation along the longitudinal axis was increased to from $\pm$1.5 to $\pm$2.5 cm. With this updated region of uncertainty, the simulated axial and longitudinal profiles (figures 12(a) and (b)) present better agreement with the experimental data.

When comparing both measurement series, S1 exhibited a $\approx$0.5 °C higher maximum temperature than S2 for the same 1000 W forward power. This could be due to uncertainties in delivered power and phase, which for the Sigma-60 system can reach up to 10% and 10°, respectively, as specified by the manufacturer. Increasing the simulated input power can reproduce the elevated peak temperature but leads to a mismatch in axial profiles because the central temperature rises uniformly, whereas the observed temperature increase appears to result from a combined effect of mapping errors and power-delivery inaccuracies. These findings underscore that agreement in peak temperature can be misleading in QA assessments and that spatial temperature profiles provide a more robust basis for evaluating system performance and model validity.

Another potential source of positional uncertainty is phantom misalignment within the applicator. Despite the use of custom-made supports to stabilize the phantom along the *x*- and *y*-axes, final alignment along the *z*-axis relied on manual, ruler-based positioning relative to the applicator bore. Assuming a plausible maximum displacement of ± 1 cm along *x*, *y*, and *z*, simulations in figures 12(c) and (d) showed that axial temperature profiles are highly sensitive to shifts of this magnitude, particularly near the phantom boundaries, which could explain localized hotspots clearly visible on the right side of the axial profiles (figure 12(c)). However, if the hotspots at the phantom wall were solely caused by phantom misalignment, the entire profile, including the center, would shift in the same way. Since this is not observed, additional unmodeled effects are likely contributing to the unexpected temperature deviations or hotspots in the measurements. These may include water-bolus flow dynamics, antenna mismatch, air gaps between the bolus and phantom, heterogeneous initial temperatures due to prior QA heating experiments, or increased EM crosstalk between antennas (Leybovich *et al* 1990). Nevertheless, there is still good overall agreement between simulations and measurements, particularly in the central region, indicating that the dominant heating behavior is adequately captured by the model. Accurate modeling of bolus-phantom boundary phenomena remains challenging, and dedicated investigations of individual additional sources of uncertainty are therefore warranted. In the clinical setting, however, their clinical impact is generally limited due to the cooling effect of the water bolus and the improved RF matching between the bolus and human tissue (Kok and Crezee 2025).

From a QA perspective, these results indicate that thermal mapping inaccuracies are the dominant source of uncertainty, while catheter displacement and phantom misalignment have a smaller influence. Current QA guidelines report acceptance thresholds for the main quality indicators, which were derived from a previous multi-institutional investigation using similar phantom setups (De Lazzari *et al* 2024). In particular, this study established that a focus localization deviation of up to 2 cm between the expected and experimentally measured focus position can be considered acceptable when using thermal mapping-based thermometry. The deviations observed in the present study fall within this recommended tolerance. Consequently, QA results should be interpreted relative to these predefined thresholds rather than in terms of absolute measurement differences. If the experimentally determined parameters deviate substantially beyond these limits, this would indicate that the system performance is outside the expected QA tolerance and may signal potential system malfunction or steering inaccuracies.

Due to time constraints in clinical practice, regular QA assessments are typically limited to single measurement sessions, which underscores the relevance of time-efficient and robust QA procedures. To minimize unpredictable thermal mapping malfunctions, which are difficult to correct retrospectively, the following strategies can be derived from the experience gained in the present work. First, thermal-mapping probe positioning should be calibrated regularly. This step is not always part of routine practice, yet inaccuracies in probe positioning can substantially affect DHT delivery, either by underestimating tumor temperatures or by missing localized hotspots along portions of the mapping path affected by probe drag. Second, with currently available systems, the simultaneous use of four or more mapping probes can markedly increase uncertainty in probe positioning. Tests performed during this study and corroborated by independent observations across multiple clinical centers indicate improved repeatability when limiting thermal mapping to two or three probes at a time. This limitation is not institution-specific and is consistent with reports from several centers within the ESHO community. A likely contributing factor is that the probes are mechanically driven simultaneously, such that variations in cable tension can translate into differential probe drag. A potential mitigation strategy that could be considered in future system designs is the use of independently actuated probe motors, allowing individual tuning of cable tension to improve mapping consistency. An alternative solution would be the use of multi-sensor probes with a 20 cm mapping length and 1–2 cm sensor spacing. However, implementing such probes with fiber-optic sensors remains cost-prohibitive. Multi-sensor thermocouple-based probes, which are routinely used in other systems (Bakker *et al* 2022b), represent a more economical option but require additional consideration due to the presence of metallic components that may introduce measurement artifacts in RF fields. Third, phantom positioning should be verified both before and after bolus filling, as buoyancy effects can introduce positional displacements that are otherwise difficult to detect and may affect the measured temperature profiles. Fourth, for routine QA purposes, emphasis should be placed on the reproducibility and symmetry of temperature profiles for a central focus, rather than on absolute peak temperature values, which are more sensitive to experimental uncertainties. Finally, documenting probe configuration, calibration status, and phantom positioning at the time of each QA measurement is essential to ensure traceability and to enable meaningful longitudinal comparisons. These recommendations are intended to improve the robustness of routine QA measurements and are not meant to replace comprehensive patient-specific thermometry or treatment planning.

4.4. Study limitations

The primary focus of this work was to analyze the effect of different parameters on the observable thermal profiles in a homogeneous phantom for QA procedures in DHT. Our analysis concentrated mainly on the geometric characteristics and dielectric properties of the gel phantom. The analysis of varying thermal properties was not included in this study nor was the temperature-dependence of both thermal and dielectric properties. However, we expect the effect of elevated temperature on thermal properties rather limited as suggests preliminary, yet unpublished, data. In particular, deviations of less than 5% in the thermal properties were observed when increasing the phantom temperature from 22 °C to 45 °C.

In this study, we carried out multiple QA experiments in a clinical DHT system, consistently applying a 10 min heat pulse focused on the phantom center with coordinates (0,0,0) cm. No other phase steering evaluations were performed. The aim was to provide a robust experimental verification of the numerical modeling under current practical constraints, including limitations in thermal acquisition and long repetition times. In addition, the central focus is clinically relevant because many centers, when feasible, align the tumor with the applicator center to minimize phase steering, as steering accuracy decreases with increasing distance from the center. Accordingly, this makes the (0,0,0) cm target represents a common and relevant QA configuration.

Given the tight clinical schedule, maintaining the experimental setup unchanged over time was challenging, and repeated phantom removal and repositioning likely introduced additional positional uncertainties. To acquire representative data, we prioritized repetitions at the central focus over QA experiments with different off-center foci, even if it does not capture potential variations in heating patterns that may occur when the focus is steered away from the applicator center. Experimental verification of off-center steering would therefore provide additional insights. However, acquiring them would have required many additional experimental sessions, which, in a busy clinical practice, would have taken several weeks. Future studies could extend the validation framework presented here to include such configurations. We therefore prioritized repeated measurements at the central focus over experiments with multiple off-center foci. Although measurements at off-center targets would have strengthened the verification, acquiring them would have required a substantial number of additional experimental sessions, which would have taken several weeks or months in a busy clinical practice. Recent work by Kok *et al* (2025) reflects a similar constraint: although different off-center foci were analyzed, the study was limited to *in silico* simulations without direct experimental verification. Their analysis of Sigma-60 data relied on comparisons with literature data, showing reasonable agreement at the central focus but larger deviations toward the phantom boundaries.

The proposed phantom also has important limitations that must be considered when interpreting QA results. Due to its homogeneous geometry and uniform dielectric and thermal properties, the phantom cannot reproduce the anatomical heterogeneity, perfusion effects, or systemic thermal responses present in patients. Consequently, measurements obtained with this phantom cannot be used to predict patient-specific temperature distributions, hotspot formation, clinical safety, or treatment effectiveness. Passing phantom-based QA verification therefore should not be interpreted as evidence of clinical safety or therapeutic efficacy. Instead, the phantom is intended solely for controlled evaluation of system performance, including verification of heating capability, focusing accuracy, and reproducibility of energy delivery. Clinical safety and effectiveness must be assessed through complementary methods, including patient-specific treatment planning, heterogeneous or anatomically realistic phantoms, comprehensive thermometry during treatment, and ultimately clinical evaluation.

5. Conclusions

This study provides an evidence-based framework for designing QA phantoms suitable for temperature-based performance assessment of DHT applicators, in line with the upcoming 2026 ESHO QA guidelines. Through systematic evaluation of geometric and material parameters, we identified phantom dimensions and dielectric properties that balance clinical relevance with practical constraints for routine QA. Dielectric properties were shown to exert a dominant influence on temperature distributions, underscoring the need for careful material selection in phantom fabrication.

Phantom-based temperature profiles were consistent with patient-model trends in the focal region, supporting the clinical relevance of the proposed QA configuration. Moreover, simulations were corroborated by repeated experimental measurements, showing good agreement at the focus, while increased

variability was observed toward the phantom boundaries. We further quantified how uncertainties in thermal mapping and phantom positioning affect the comparison between simulated and measured temperature profiles. To capture typical mapping uncertainty, average positioning errors of ± 1.5 cm along the probe axis and ± 0.5 cm perpendicular to the axis should be accounted for. However, these margins do not capture position errors driven by pronounced probe drag, which can create an artificial plateau in the thermal profile; this risk can be mitigated by limiting the number of mapping probes used simultaneously and by regular calibration of mapping position.

Overall, the present work strengthens the technical foundation for standardized temperature-based QA in DHT by linking phantom design choices to measurable performance metrics and by providing practical guidance for improving the robustness and interpretability of thermal-mapping measurements.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary information files).

ORCID iDs

Mattia De Lazzari 0000-0002-9736-5457
Hana Dobšíček Trefná 0000-0001-6025-0819
Carolina C Seabra 0009-0002-0767-8968
Patrick V Granton 0000-0002-2635-2894
Sergio Curto 0000-0002-3073-1117
Dario B Rodrigues 0000-0001-6805-5989

References

Bakker A *et al* 2022a Post-operative re-irradiation with hyperthermia in locoregional breast cancer recurrence: temperature matters *Rad. Oncol.* **167** 149–57

Bakker A, van der Zee J, van Tienhoven G, Kok H P, Rasch C R N and Crezee H 2019 Temperature and thermal dose during radiotherapy and hyperthermia for recurrent breast cancer are related to clinical outcome and thermal toxicity: a systematic review *Int. J. Hyperthermia* **36** 1023–38

Bakker A, Zweije R, Kok H P, Stalpers L J A, Westerveld G H, Hinnen K A, van Tienhoven G, Kolff M W and Crezee H 2022b Comparison of the clinical performance of a hybrid Alba 4D and the AMC-4 locoregional hyperthermia systems *Int. J. Hyperthermia* **39** 1408–14

Bellizzi G G, Sumser K, VilasBoas-Ribeiro I, Curto S, Drizdal T, van Rhoon G C, Franckena M and Paulides M M 2020 Standardization of patient modeling in hyperthermia simulation studies: introducing the Erasmus Virtual Patient Repository *Int. J. Hyperthermia* **37** 608–16

Bruggmoser G *et al* 2011 Quality assurance for clinical studies in regional deep hyperthermia *Strahlenther. Onkol.* **187** 605

Curto S *et al* 2019 Quantitative, multi-institutional evaluation of MR thermometry accuracy for deep-pelvic MR-hyperthermia systems operating in multi-vendor MR-systems using a new anthropomorphic phantom *Cancers* **11** 1709

Datta N R, Ordóñez S G, Gaipl U S, Paulides M M, Crezee H, Gellermann J, Marder D, Puric E and Bodis S 2015 Local hyperthermia combined with radiotherapy and-/or chemotherapy: recent advances and promises for the future *Cancer Treat. Rev.* **41** 742–53

Datta N R, Puric E, Klingbiel D, Gomez S and Bodis S 2016 Hyperthermia and radiation therapy in locoregional recurrent breast cancers: a systematic review and meta-analysis *Int. J. Radiat. Oncol. Biol. Phys.* **94** 1073–87

De Lazzari M, Carrapiço-Seabra C, Marder D, van Rhoon G C, Curto S and Dobšíček Trefná H 2024 Toward enhanced quality assurance guidelines for deep hyperthermia devices: a multi-institution study *Int. J. Hyperthermia* **41** 2436005

Dobšíček Trefná H *et al* 2025 Submitted to International Journal of Hyperthermia *Quality assurance guidelines for performance assessment of phased-array deep hyperthermia therapy systems*

Farina L, Sumser K, van Rhoon G and Curto S 2020 Thermal characterization of phantoms used for quality assurance of deep hyperthermia systems *Sensors* **20** 4549

Franckena M, Fatehi D, Bruijne M D, Canters R A M, Norden Y V, Mens J W, Rhoon G C V and Zee J V D 2009 Hyperthermia dose-effect relationship in 420 patients with cervical cancer treated with combined radiotherapy and hyperthermia *Eur. J. Cancer* **45** 1969–78

Hasgall PA *et al* 2022 *IT'IS database for thermal and electromagnetic parameters of biological tissues*

Horsman M and Overgaard J 2007 Hyperthermia: a potent enhancer of radiotherapy *Clin. Oncol.* **19** 418–26

Issels R D *et al* 2010 Neo-adjuvant chemotherapy alone or with regional hyperthermia for localised high-risk soft-tissue sarcoma: a randomised phase 3 multicentre study *Lancet Oncol.* **11** 561–70

James B J and Sullivan D M 1992 Creation of three-dimensional patient models for hyperthermia treatment planning *IEEE Trans. Biomed. Eng.* **39** 238–42

Joines W T, Zhang Y, Li C and Jirtle R L 1994 The measured electrical properties of normal and malignant human tissues from 50 to 900 MHz *Med. Phys.* **21** 547–50

Kok H and Crezee J 2025 Validation of the implementation of phased-array heating systems in Plan2Heat *Strahlenther. Onkol.* **201** 135–50

Laakso I *et al* 2026 Roadmap towards personalized approaches and safety considerations in non-ionizing radiation: from dosimetry to therapeutic and diagnostic applications *Phys. Med. Biol.* **71** 03RM01
Lagendijk J *et al* 1998 ESHO quality assurance guidelines for regional hyperthermia *Int. J. Hyperthermia* **14** 125–33
Leopold K A *et al* 1992 Relationships among tumor temperature, treatment time, and histopathological outcome using preoperative hyperthermia with radiation in soft tissue sarcomas *Int. J. Radiat. Oncol. Biol. Phys.* **22** 989–98
Leybovich L, Myerson R J, Emami B and Straube W L 1990 Evaluation of the sigma—60 applicator for regional heating in terms of scattering parameters *Int. J. Radiat. Oncol. Biol. Phys.* **19** 203
Lutgens L *et al* 2010 Combined use of hyperthermia and (chemo) radiation therapy for treating locally advanced cervix carcinoma *Cochrane Database Syst. Rev.* **1**
Oei A, Kok H P, Oei S B, Horsman M R, Stalpers L J A, Franken N A P and Crezee J 2020 Molecular and biological rationale of hyperthermia as radio-and chemosensitizer *Adv. Drug Delivery Rev.* **163** 84–97
Paulides M M, Rodrigues D B, Bellizzi G G, Sumser K, Curto S, Neufeld E, Montanaro H, Kok H P and Dobsicek Trefna H 2021 ESHO benchmarks for computational modeling and optimization in hyperthermia therapy *Int. J. Hyperthermia* **38** 1425–42
Rijnen Z, Bakker J F, Canters R A M, Togni P, Verduijn G M, Levendag P C, Van Rhoon G C and Paulides M M 2013 Clinical integration of software tool VEDO for adaptive and quantitative application of phased array hyperthermia in the head and neck *Int. J. Hyperthermia* **29** 181–93
Schneider C, Olmi R and Van Dijk J 1995 Phantom design: applicability and physical properties *Thermoradiotherapy and Thermochemotherapy: Biology, Physiology, Physics* (Springer) pp 381–97
Schneider C, Van Dijk J D P, De Leeuw A A C, Wust P and Baumhoer W 1994 Quality assurance in various radiative hyperthermia systems applying a phantom with LED matrix *Int. J. Hyperthermia* **10** 733–47
Schneider C and Van Dijk J 1991 Visualization by a matrix of light-emitting diodes of interference effects from a radiative four-applicator hyperthermia system *Int. J. Hyperthermia* **7** 355–66
Silva N P, Bottiglieri A, Conceição R C, O'Halloran M and Farina L 2020 Characterisation of *ex vivo* liver thermal properties for electromagnetic-based hyperthermic therapies *Sensors* **20** 3004
Van der Gaag M, De Bruijne M, Samaras T, Van Der Zee J and Van Rhoon G C 2006 Development of a guideline for the water bolus temperature in superficial hyperthermia *Int. J. Hyperthermia* **22** 637–56
Van Driel W J *et al* 2018 Hyperthermic intraperitoneal chemotherapy in ovarian cancer *New Engl. J. Med.* **378** 230–40
Willner A *et al* 2021 Neoadjuvant concurrent chemoradiotherapy with and without hyperthermia in retroperitoneal sarcomas: feasibility, efficacy, toxicity, and long-term outcome *Strahlenther. Onkol.* **197** 1063–71
Zanoli M and Dobšíček Trefná H 2022 The hot-to-cold spot quotient for SAR-based treatment planning in deep microwave hyperthermia *Int. J. Hyperthermia* **39** 1421–39
Zanoli M, Ek E and Dobšíček Trefná H 2023 Antenna arrangement in UWB helmet brain applicators for deep microwave hyperthermia *Cancers* **15** 1447