Title: Multi-institutional dosimetric delivery assessment of intracranial stereotactic radiosurgery on different treatment platforms

Keywords: Radiosurgery, Dosimetry, End-to-end, Audit, Anthropomorphic Phantom, Alanine, Radiochromic film

Highlights:

- A multi-institutional end-to-end assessment of radiosurgery dosimetric delivery was performed for 33 plans in 30 participating centres using a variety of treatment platforms.
- The comparison has highlighted the dosimetric consistency achievable with different delivery platforms.
- The need for standardisation in intracranial stereotactic radiosurgery is highlighted.

Abstract:

Background and Purpose:
Assessment of dosimetric accuracy of radiosurgery on different treatment platforms.
Material and Methods:
Thirty-three single fraction treatment plans were assessed at thirty centres using an anthropomorphic head phantom with target and brainstem structures. The target being a single irregular shaped target, ~8cc, 10 mm from the brainstem. The phantom was “immobilised”, scanned, planned and treated following the local protocols. EBT-XD films and alanine pellets were used to measure absolute dose, inside both the target and the brainstem, and compared with TPS predicted dose distributions.
Results:
PTV alanine measurements from gantry-based linacs showed a median percentage difference to the TPS of 0.65%. Cyberknife (CK) had the highest median difference of 2.3% in comparison to the other platforms. GammaKnife (GK) showed the smallest median of 0.3%. Similar trends were observed in the OAR with alanine measurements showing median percentage differences of1.1%, 2.0% and 0.4%, for gantry-based linacs, CK and GK respectively. All platforms showed comparable gamma passing rates between axial and sagittal films.
Conclusions:
This comparison has highlighted the dosimetric variation between measured and TPS calculated dose for each delivery platforms.. The results suggest that clinically acceptable agreement with the predicted dose distributions is achievable by all treatment delivery systems.

1. Introduction

Stereotactic radiosurgery (SRS), was first developed in the 1950s [1] and has since evolved substantially. There are now several manufacturers that offer commercial solutions for delivering SRS and such treatments may be delivered by a Gamma Knife (GK) unit, a CyberKnife (CK) or a gantry-based linear accelerator (linac) system with stereotactic capabilities.

All radiotherapy practices should be subjected to appropriate quality assurance procedures, including regular quality control testing and independent external dosimetry audit [2,3], to minimize potential errors in treatment delivery that can lead to clinical complications [4]. This is especially the case for SRS where a very high dose is delivered in only a single fraction, meaning an error cannot be mitigated in a subsequent fraction.

The multiple platforms which can be used for SRS are very different in terms of their technologies and techniques [5]. Furthermore, they have been, and are being, used very differently in terms of their dosimetric and clinical practices [6]. The categorisation of the various systems into Gamma Knife (GK) units, CyberKnife (CK) units or gantry-based linear accelerator (linac) systems with stereotactic capabilities is justified according to their broadly similar dosimetric and clinical practices [6]. The latter category includes systems which use cones or tertiary microMLCs and also includes tomotherapy units, which use a fixed ring gantry and have no non-coplanar delivery capability.

This national study was undertaken to support an initiative in the UK to regulate the provision of cranial SRS services [7–9]. The participating centres were audited in an end-to-end test, incorporating local clinical procedures for immobilisation devices, CT-scanning, target contouring, treatment planning and treatment delivery. We evaluated the agreement between planned and delivered dose for each of the audited systems. It is acknowledged that the audited systems cover a wide range of delivery and ancillary systems, however such a study has the potential to benchmark what can be achieved with intracranial stereotactic surgery systems and therefore contribute data for use in setting tolerances for clinical trials and future external audits. Furthermore, the variations in clinical practice observed between these different platforms have been evaluated to assess the need for standardisation.

2. Materials and Methods

A comprehensive end-to-end test was developed based on an anthropomorphic head phantom, STEEV (Stereotactic End-to-End Verification, CIRS, Norfolk, VI, USA). The phantom was adapted to contain a single irregularly shaped target (~8cc), 10mm anterior to the brainstem, for treatment by the audited centre following their local protocol for brain metastasis to be treated in a single fraction [10]. A Computed Tomography (CT) scan of the phantom was sent in advance to the participating centres in DICOM format for volume contouring and pre-planning. This was followed by a visit to the centre where the phantom was scanned following the local protocol (CT only), with eight dummy alanine pellets and two dummy pieces of EBT-XD Gafchromic film (Ashland ISP Advanced Materials, NJ, USA) placed inside the phantom to mimic the detector positions. These were subsequently replaced with the real detectors prior to phantom treatment irradiation. A graphical representation showing the target (PTV), brainstem (OAR) and detector positions is included in Figure 1. The two CT scans were co-registered, and the pre-plan transferred to the local scan and finalised.

Thirty centres participated in the audit. In three centres two treatment platforms were assessed and hence thirty-three single fraction treatment plans were generated and delivered to the phantom. Table 1 provides details for each platform and plan that participated in the audit. Further details of the auditing protocol followed are given in Figure 2.

2.1. Reference beam output measurements

A calibrated PTW 31010 semiflex ionisation chamber (0.125cc), traceable to a graphite calorimeter primary standard at the National Physical Laboratory (NPL, Teddington, UK) was used in all centres to perform reference beam output measurements. The chamber was placed in an auditor provided water-equivalent plastic material, using the centre's reference conditions, to measure the output in the machine specific reference field [11]. All such auditor measurements were corrected for temperature and pressure. Temperature and pressure were measured using independent auditor equipment. The measurements were performed in machine-specific reference conditions and deviations from agreement were expressed relative to the TPS-calculated value for dose in a geometric phantom.
The reference beam output measurements were used to apply corrections to the alanine measurements of the clinical plan, to compensate for any daily output variation.

2.2. Alanine measurements

Two sets of four alanine pellets were irradiated in each centre: one in the target and one in the brainstem. The phantom temperature before and after each measurement was recorded and used to apply a temperature correction factor. All pellets were returned to NPL and were processed within 1 month of the audit visit, ensuring negligible fading [12]. The measured doses were compared to the mean dose calculated by the TPS for each pellet. Additionally, in order to account for positional uncertainties that can lead to large percentage differences between measured and predicted doses, mean doses for each pellet stack were compared with the TPS prediction for the complete stack.

The percentage difference between alanine and TPS for the OAR was normalised to the 12 Gy dose level, being the nominal brainstem tolerance dose used by many centres. This enabled a dosimetric comparison whilst assessing the plan quality in terms of overdose to the OAR.

Uncertainty on the alanine readout was taken to be 1.4% (k=2) [12]. The deviations from agreement were expressed relative to the TPS-calculated value.

Alanine was considered as the primary detector for the audit measurements, as it has proven its efficacy in small fields and is provided by an established dosimetry service based at NPL [12].

2.3. Film measurements

The film response was calibrated in a conventional manner with ten EBT-XD film pieces exposed in the range of 0-40 Gy. The films were irradiated in a 10 cm × 10 cm field at 5 cm depth in water-equivalent plastic, with a nominal 6 MV beam. The scanned pixel value as a function of dose was determined as the average pixel value in a 4 × 4 cm region centred on the beam axis. Images were converted to dose maps using FilmQAPro® software (Ashland ISP Advanced Materials, NJ, USA) using the red-green-blue triple-channel dosimetry algorithm [13].

Two films (one sagittal and one axial) were placed inside the phantom and simultaneously irradiated at each centre. All films were returned to NPL and scanned at least 72 hours after exposure following an established film dosimetry protocol [14,15] and good film dosimetry practice [16,17] with the analysis performed on FilmQAPro. Film-dose linear scaling was applied using reference films at zero dose and 80% of the maximum anticipated dose from the treatment plan, which were scanned simultaneously with the test films. This approach mitigates the effects of post-exposure darkening and variations of the scanner response, and stabilizes the calibration (forced into agreement) at the reference dose levels [18]. The regions of interest used for the analysis were a 6 x 5 cm rectangle for the axial films and a 7 x 4 cm rectangle for the sagittal films. The measured dose distributions were matched to TPS dose distributions using the optimum shift algorithm, where small shifts were applied (<1mm and <1^{o}) to optimise the gamma passing rate..

Gamma passing rates [19] were collected for a range of criteria (global and local) using the red colour channel with triple-channel-correction and with film-dose linear scaling corrections applied. The criteria selected were based on the following: data for 3% dose difference and 2mm distance (3%/2mm) were collected as this criterion was commonly used by the participating linac users and 5%/1mm as this is more suitable for plans with steep gradients. Data for 2%/2mm was also analysed as recommended by AAPM TG 135 [20]. A minimum cut-off threshold of 2 Gy was applied to remove low

dose areas of higher uncertainty from the analysis. This absolute value was used as an alternative to a relative value, due to the large variation of prescription doses and maximum doses delivered by the audit participants.

2.4 Statistical analysis

A Kruskal-Wallis test was performed to detect if there were any significant differences in median values between the three platforms, followed by the Bonferroni-type multiple comparison to establish the hierarchy/significant differences between different pairs.

3. Results

Reference output measurements were performed independently by both the host centre and the auditors using their respective equipment and these are presented in Figure 3. All measurements from the auditors and the centres were within ±2.4% of the dose calculated by the TPS, ranging from -1.0% to +2.4%. The mean difference between auditor measurement and host centre measurement (audit/host) was +0.5%, with a maximum difference of +1.2%. The gantry based linac group spread in the output measurements was 3.2%, ranging from -0.8% to +2.4%, with a median of 0.35%. CKs had a spread of 1.8%, ranging from -1.0% to +0.8%, with a median of 0.85% and GKs had a of spread, 1.6%, ranging from -0.9% to +0.7%, with a median of 0.0%. No statistical differences were seen between the platforms.

The differences between the measured and calculated (measured/calculated) doses for the stack of 4 alanine pellets ranged from -1.3% to +4.0% for the PTV (Figure 4A). The gantry based linacs showed a spread in percentage difference of 5.2% (from -1.3% to +3.9%) with a median of 0.65%. GKs showed a spread of 2.4% (from -0.8% to +1.5%), with a median percentage difference of+0.3%. CK measurements had a spread of 2.6% (from +1.4% to +4%), with a median of +2.3%, which in comparison with the gantry based linac and GK groups were statistically higher with p values of 0.045 and 0.039 respectively. Individual pellet measurements in the target showed differences of up to 14% when compared to the TPS-predicted mean dose in their individual contoured pellet structure.

Similar trends were observed in the comparison of OAR alanine pellet measurements to those observed in target alanine pellet measurements, ranging from -1.1% to +4.3% (Figure 4B). Gantry-based linac measurements showed a spread at 4.6% ranging from -1% to +3.6% with a median of +1.2%. CK measurements ranged between 0.0% to +1.9% with a median of +1.1%. GKs had a spread of 2.0% ranging from -1.1% to +0.9% and a median of +0.4%. Here the only significance was in the gantry-based linac group being higher than the GK group with a p value of 0.047.

TPS-predicted doses for the alanine pellets in the OAR ranged from 0.3 Gy up to 7.5 Gy. Figure 4 shows the percentage difference between the mean dose measured by OAR alanine pellets with the TPS-predicted mean dose, normalised to 12.0 Gy.

The results for axial and sagittal films (example shown in Supplementary Figure 1) showed consistency in the gamma passing rates achieved. The gamma passing rates for 3% / 2mm local gamma and 5% / 1mm global gamma are shown in Figure 5.

Higher passing rates for all films were observed for global than for local gamma criteria. For the 3% / 2 mm local gamma criterion, all but two films showed passing rates above 75%. The CK, GK and gantry-based linac groups had median passing rates of 88.7, 92.8 and 85.5 respectively, showing no statistical differences between them. For the 5% / 1 mm global gamma criterion, all but 3 films showed passing rates above 90% (see Figure 5). Here the CK, GK and gantry-based linac groups had median passing rates of 99.3, 98.4, 98.3 respectively, again showing no statistical differences between them. For the 2% / 2 mm global gamma criterion, all but 3 films showed passing rates above 90% (see Supplementary Figure 2). Here the CK, GK and gantry-based linac groups had median passing rates of 99.3, 99.0, 96.6 respectively, with a p value of 0.028 between the GK and the gantry-based linac groups, all other comparisons being non-significant.

When the regions of interest used for the gamma analysis were reduced to smaller areas to include the target region only, passing rates improved substantially for all centres, showing very good agreement (>95%) between TPS-predicted and delivered

dose distributions. The majority of failed pixels for all films analysed were found to be outside the target, between the 2 Gy (threshold level) and the 12 Gy isodose line.

4. Discussion

The reference output measurements performed in the standard conditions for each platform at all centres were within ±2.4% of the expected dose, well within the ±3% recommendation of IPEM report 81 [21]. Moreover, the results seen in this study are consistent with the results reported by the NPL over 20 years of reference audits [22]. The differences from expected dose seen in the GK and CK groups were comparable, and much smaller than those seen in gantry-based linacs. This could be related to these platforms having more predictable output in reference conditions, due to having a simpler design. The ^{60}Co sources in the GK have a predictable decay, which is reflected in the good agreement seen in these measurements and the tighter tolerances of 1.0 - 1.5% deviation used by GK centres. The compact single energy linear accelerator of the CK may be the reason for the reduced fluctuations observed in output measurements, compared to multi-energy and multi-modality conventional gantry-based linacs which comprise the majority of units within this category. The systematic difference between the auditor and the centres (+0.5%) may be partly explained by the instruments used. While the auditors used a 0.125cc ionisation chamber for all measurements, the centres used a range of different detectors, which in most cases had a larger sensitive volume. Therefore, a small degree of volume averaging can be attributed to the difference observed, especially since 15 out of 33 participating platforms used Flattening Filter Free (FFF) type beams. However, further investigation needs to be performed to fully explain this difference.

Overall, good agreement was observed between alanine and TPS, with three centres falling outside (greater than) two standard deviations of the mean (two centres in the target dose measurements and one in the OAR measurements). Although there were some statistical differences between the groups these had p values which were only just less than 0.05 and were not consistent across the measurement methods and PTV/OAR, hence no strong statistical conclusions can be drawn about each platform. The gantry-based linac group was seen to have the largest spread in percentage

differences compared to CK and GK, but with a relatively good overall agreement to the TPS. The reasons behind this spread are most likely attributed to the diversity of techniques, platforms, beam energies, TPSs, calculation algorithms, clinician preferences and influences from the local radiotherapy practices. In comparison, the CK and GK groups had almost identical settings and practices within their subgroups. It is also possible that the variations in the commissioning methodologies used for gantry-based linacs, especially with regards to the dosimeters used for the measurement of small radiation fields, has an impact on the accuracy of dose calculation. Although numerous studies have been conducted on appropriate detectors for small field applications [23–28], there has been a lack of international guidelines until the recently published IAEA TRS 483 [29]. It is expected that these new guidelines will improve the standardisation of commissioning methodologies.

CK measurements in the target, showed that the TPS with Raytracing calculation algorithm under-estimated the dose in all four centres visited. This finding is in agreement with another study utilising the same alanine service for target dose measurements in CK plans [30].

All GK centres used a TMR10 algorithm that does not account for density inhomogeneities and assumes water density within a CT-generated or depth-helmet measured skull contour. A recent study investigating the GK convolution algorithm (employs density corrections) in comparison to the TMR10, showed a 6% difference between the two where 1.5% of this was attributed to depth helmet measurements [31]. Our alanine measurements showed good agreement with the TPS and suggest that these sources of error do not contribute significantly to dosimetric inaccuracies. Further work investigating this convolution algorithm may be required to evaluate its accuracy before it is used clinically.

Alanine measurements in the OAR were performed along a steep dose gradient where positional uncertainties may be expressed as large dose differences. Doses to the OAR pellets ranged by up to an order of magnitude, caused by individual planning priorities and protocols used by participating centres. As some centres delivered very low doses to this region the reported doses were normalised to 12 Gy (a nominal brainstem tolerance dose value used by many centres) in order to provide a useful measure of relative accuracy to the centres with different approaches. There was also

a higher uncertainty in the lower dose measurements due to lower signal to noise ratios in the alanine readout.

The gamma passing rates showed clinically acceptable agreement between the film-measured dose and the treatment planning system calculated dose distributions for both sagittal and axial films [32]. All treatment modalities showed comparable variations in passing rates between the centres assessed and the passing rates alone do not suggest significant differences between the different platforms. Other studies have suggested that gamma index analysis is non-ideal for direct comparison in multi-institution assessments due to inherent differences in the dose distributions, particularly in dose gradient and maximum doses [32–34]. Local Gamma criteria may favour linac centres in which the dose gradient could be less steep than GK and CK, and Global Gamma criteria may favour GK centres in which the maximum (normalisation) dose is higher. Whilst the methodology for film analysis employed in this study was designed to diminish sensitivities to different dose distributions it is impossible to achieve this with gamma index analysis. Gamma passing rates are also sensitive to the position of the film relative to the dose plan, the position of the region of interest used for the analysis and the 2 Gy threshold levels applied. An analysis method that is less sensitive to these dose distribution differences could be preferable, enabling a more reliable direct comparison between competing plans. The use of dose-plane-histograms (analogous to dose-volume histograms (DVH) in a single plane) may provide a more clinically relevant analysis [35,36]. Despite its pitfalls, the gamma analysis method used, enabled quantification of the dose shaping abilities of all SRS platforms active in the UK. The results showed clinically acceptable dosimetric performance by all platforms, although noticeable dosimetric differences were apparent outside the target volume, which are unlikely to be clinically relevant. These dosimetric inaccuracies, seen in most centres, are related to the TPS's limitations in simulating out of field doses and typically resulted in underestimation of doses to the OAR, as found by other studies [37]. Another study conducting film-based end-to-end tests in CK plans recorded higher gamma passing rates (>90%) for the criterion of 2%-2mm local gamma [38], compared to the median passing rates seen in this study of 76.8, 81.0 and 75.1% for CK, GK and gantry-based linacs respectively. The differences seen are explained by the higher 50%

dose threshold used in that study that excludes the low doses included in our film analysis.

While other multi-platform assessments have been conducted on stereotactic applications [30,39,40] only a few have been performed specifically for SRS [41].

Conclusion

This study is novel in the diversity of treatment platforms included and the advanced dosimetry methods employed. During this study, thirty-three treatment plans were generated, all for the same realistic patient scenario of a single metastatic lesion located anterior to the brainstem. The approaches adopted by the participants to treat the presented indication differed in many aspects (see Table 1). Some of these differences, with respect to the equipment, software and delivery techniques used, were previously identified [6]. Aside from these, some subtle differences were observed in the accuracy of the measured dose distributions. However, the most influential and clinically relevant variation observed in the protocols assessed, was found in prescription practices (Table 1), highlighting the potential need for standardisation.

This study assessed the dosimetric accuracy achieved using the thirty-three participating platforms. Although a statistical analysis was performed, there were small numbers in each group and a more robust analysis can be performed with a larger population. However, this was a national study and a larger study can only be realised at a multi-national level. A larger study may also reveal differences in the individual approaches followed in each centre. Another limitation of this work is the lack of incorporation of MRI in the end-to-end assessment, which is an integral step in intracranial SRS. As mentioned previously, there are also limitations with the use of gamma analysis in this setting. Future studies of this nature must develop novel methodologies to enable more meaningful and clinically relevant comparisons, such as 3D dosimetry and improvements in the analysis of dose distributions moving towards a DVH-based assessment.

With the recent rise of gantry based linac SRS [6] it is essential to incorporate all SRS platforms in dosimetric studies and clinical trials, in order to reach consensus in

several matters. SRS treatments have been traditionally split between neurosurgery-led single session SRS and multisession oncology-led practices. However, the promising outcomes seen in hypofractionated SRS [42] and staged-SRS [43] has brought the two faculties closer. In the future stereotactic trials should consider the diversity of platforms that may be used, in particular for issues such as prescription dose, prescription isodose, delivery capability, and how dosimetric differences may be assessed [44].

Currently, the assessment of dosimetric deliveries in a multi-platform, multi-centre SRS setting with individual planning priorities remains a challenge. However, independent dosimetric assessments as presented in this study are important interventions which have a crucial role in ensuring accurate dose delivery to patients. Moreover, when data from multiple centres is pooled together, it enables participants to benchmark their services against the rest of their community, assess their safety, evaluate their practices and consider improvements to their service. As stated the basis of this audit was to support an initiative in the UK to regulate the provision of cranial SRS services. Consistent dosimetry has been recognised as essential in the evaluation of outcomes from those centres commissioned to provide such clinical services. Overall, there was a smaller spread of data seen in the CK and GK groups, however each had some statistically significant differences with the other platforms. Future multicentre SRS studies may benefit from some standardisation and consensus of practice.

5. Acknowledgments

Funding for PhD studies and partial travel costs - Engineering & Physical Sciences Research Council (EPSRCEP/J500094), Technical support (Ilias Billas, Michael Homer, Ana Subiel, Simon Duane, Sebastian Galer), Alanine analysis (Gavin Cox, David Crossley), Audit support (Elizabeth Miles, Rada Zotova), Pilot audit support (Martyn Gilmore, Phillip Cooper, Kelvin Hiscoke).

6. References

[1] Leksell L. The stereotaxic method and radiosurgery of the brain. Acta Chir Scand 1951;102:316–9.

[2] Clark CH, Jornet N, Muren LP. The role of dosimetry audit in achieving high quality radiotherapy 2018. https://doi.org/10.1016/j.phro.2018.03.009.

[3] Clark CH, Ga Aird E, Bolton S, Miles EA, Nisbet A, Snaith JA, et al. Radiotherapy dosimetry audit: Three decades of improving standards and accuracy in UK clinical practice and trials. Br J Radiol 2015;88. https://doi.org/10.1259/bjr.20150251.

[4] Briggs G, Ebdon-Jackson S, Erridge SC, Graveling M, Hood S, McKenzie A, et al. Towards Safer Radiotherapy. R Coll Radiol Soc Coll Radiogr Inst Phys Eng Med Natl Patient Saf Agency Br Inst Radiol 2008:85. https://doi.org/978-1-905034-25-3.

[5] Kocher M, Wittig A, Piroth MD, Treuer H, Seegenschmiedt H, Ruge M, et al. Stereotactic radiosurgery for treatment of brain metastases: A report of the DEGRO Working Group on Stereotactic Radiotherapy. Strahlentherapie Und Onkol 2014;190:521–32. https://doi.org/10.1007/s00066-014-0648-7.

[6] Dimitriadis A, Kirkby KJ, Nisbet A, Clark CH. Current Status of Cranial Stereotactic Radiosurgery in the UK. Br J Radiol 2016;89:20150452. https://doi.org/10.1259/bjr.20150452.

[7] NHS England. Stereotactic Radiosurgery/ Stereotactic Radiotherapy Needs Assessment and Service Review Consultation Report. 2015.

[8] NHS England. Service Specifications: Stereotactic radiosurgery and stereotactic radiotherapy (Intracranial) (All ages). London: 2016.

[9] Eaton DJ, Lee J, Paddick I. Stereotactic radiosurgery for multiple brain metastases: Results of multicenter benchmark planning studies. Pract Radiat Oncol 2018;0. https://doi.org/10.1016/j.prro.2017.12.011.

[10] Dimitriadis A, Palmer AL, Thomas RAS, Nisbet A, Clark CH. Adaptation and validation of a commercial head phantom for cranial radiosurgery dosimetry end-to-end audit. Br J Radiol 2017;90:20170053. https://doi.org/10.1259/bjr.20170053.

[11] Alfonso R, Andreo P, Capote R, Huq MS, Kilby W, Kjäll P, et al. A new formalism for reference dosimetry of small and nonstandard fields. Med Phys 2008;35:5179–86. https://doi.org/10.1118/1.3005481.

[12] Sharpe PHG, Sephton JP. Therapy level alanine dosimetry at the NPL. Proc. 216th PTB Semin. Alanine Dosim. Clin. Appl. PTB-Dos-51, PTB, Braunschweig, 2006.

[13] Micke A, Lewis DF, Yu X. Multichannel film dosimetry with nonuniformity correction. Med Phys 2011;38:2523–34. https://doi.org/10.1118/1.3611636.

[14] Palmer AL, Dimitriadis A, Nisbet A, Clark CH. Evaluation of Gafchromic EBT-XD film, with comparison to EBT3 film, and application in high dose

448 radiotherapy verification. Phys Med Biol 2015;60:8741–52.
449 https://doi.org/10.1088/0031-9155/60/22/8741.
450 [15] Dimitriadis A. Assessing the dosimetric and geometric accuracy of stereotactic
451 radiosurgery. University of Surrey, 2017.
452 [16] Niroomand-Rad A, Blackwell CR, Coursey BM, Gall KP, Galvin JM,
453 McLaughlin WL, et al. Radiochromic film dosimetry: Recommendations of
454 AAPM Radiation Therapy Committee Task Group 5. Med Phys 1998;25:2093–
455 115. https://doi.org/10.1118/1.598869.
456 [17] Palmer A, Bradley D, Nisbet A. Evaluation and mitigation of potential errors in
457 radiochromic film dosimetry due to film curvature at scanning. J Appl Clin Med
458 2015;16:425–31.
459 [18] Lewis D, Micke A, Yu X, Chan MF. An efficient protocol for radiochromic film
460 dosimetry combining calibration and measurement in a single scan. Med Phys
461 2012;39:6339–50. https://doi.org/10.1118/1.4754797.
462 [19] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative
463 evaluation of dose distributions. Med Phys 1998;25:656–61.
464 https://doi.org/10.1118/1.598248.
465 [20] Dieterich S, Cavedon C, Chuang CF, Cohen AB, Garrett JA, Lee CL, et al.
466 Report of AAPM TG 135: Quality assurance for robotic radiosurgery. Med
467 Phys 2011;38:2914–36. https://doi.org/10.1118/1.3579139.
468 [21] Lillicrap SC. Physics Aspects of Quality Control in Radiotherapy (Report No.
469 81). Phys Med Biol 2000;45:815–815. https://doi.org/10.1088/0031-
470 9155/45/3/501.
471 [22] Thomas RAS, Bolt MA, Bass G, Nutbrown R, Chen T, Nisbet A, et al.
472 Radiotherapy reference dose audit in the United Kingdom by the National
473 Physical Laboratory: 20 years of consistency and improvements. Phys Imaging
474 Radiat Oncol 2017;3:21–7. https://doi.org/10.1016/j.phro.2017.07.005.
475 [23] Silvestre I, Dimitriadis A, Subiel A. Relative output factors for TPS beam data
476 acquisition with emphasis on small fields. Comparison of multiple detectors
477 and multiple approaches. Rev Latinoam Fis Medica - ISSN 2413-9904 2017;3.
478 [24] Dimitriadis A, Patallo Silvestre I, Billas I, Duane S, Nisbet A, Clark CH, et al.
479 Characterisation of a plastic scintillation detector to be used in a multicentre
480 stereotactic radiosurgery audit. Radiat Phys Chem 2017:0–1.
481 https://doi.org/10.1016/j.radphyschem.2017.02.023.
482 [25] Bassinet C, Huet C, Derreumaux S, Brunet G, Chéa M, Baumann M, et al.
483 Small fields output factors measurements and correction factors determination
484 for several detectors for a CyberKnife® and linear accelerators equipped with
485 microMLC and circular cones. Med Phys 2013;40:071725.
486 https://doi.org/10.1118/1.4811139.
487 [26] Kamio Y, Bouchard H. Correction-less dosimetry of nonstandard photon fields:
488 A new criterion to determine the usability of radiation detectors. Phys Med Biol
489 2014;59:4973–5002. https://doi.org/10.1088/0031-9155/59/17/4973.
490 [27] Russo S, Masi L, Francescon P, Frassanito MC, Fumagalli ML, Marinelli M, et
491 al. Multicenter evaluation of a synthetic single-crystal diamond detector for
492 CyberKnife small field size output factors. Phys Medica 2016.
493 https://doi.org/10.1016/j.ejmp.2016.03.005.
494 [28] Pimpinella M, Ciancaglioni I, Consorti R, Venanzio C Di, Guerra a S, Petrucci
495 a, et al. A synthetic diamond detector as transfer dosimeter for D w
496 measurements in photon beams with small field sizes. Metrologia
497 2012;49:S207–10. https://doi.org/10.1088/0026-1394/49/5/S207.

[29] Palmans H, Andreo P, Huq S, Seuntjens J. Dosimetry of small static fields used in external beam radiotherapy: An IAEA-AAPM International Code of Practice for reference and relative dose determination. Technical Report Series No. 483. Iaea Trs483 2017.
[30] Distefano G, Lee J, Jafari S, Gouldstone C, Baker C, Mayles H, et al. A national dosimetry audit for stereotactic ablative radiotherapy in lung. Radiother Oncol:406–10. https://doi.org/10.1016/j.radonc.2016.12.016.
[31] Rojas-villabona A, Kitchen N, Paddick I. Investigation of dosimetric differences between the TMR 10 and convolution algorithm for Gamma Knife stereotactic radiosurgery. J Appl Clin Med Phys 2016;17:1–13.
[32] Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Med Phys 2018;45:e53–83. https://doi.org/10.1002/mp.12810.
[33] Ezzell GA, Burmeister JW, Dogan N, LoSasso TJ, Mechalakos JG, Mihailidis D, et al. IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. Med Phys 2009;36:5359–73. https://doi.org/10.1118/1.3238104.
[34] Hussein M, Clark CHH, Nisbet A. Challenges in calculation of the gamma index in radiotherapy – Towards good practice. Phys Medica 2017;36:1–11. https://doi.org/10.1016/j.ejmp.2017.03.001.
[35] Palmer AL, Nash D, Kearton JR, Jafari SM, Muscat S. A multicentre “end to end” dosimetry audit of motion management (4DCT-defined motion envelope) in radiotherapy. Radiother Oncol 2017;125:453–8. https://doi.org/10.1016/j.radonc.2017.09.033.
[36] Sothmann T, Blanck O, Poels K, Werner R, Gauer T. Real time tracking in liver SBRT: Comparison of CyberKnife and Vero by planning structure-based γ-evaluation and dose-area-histograms. Phys Med Biol 2016;61:1677–91. https://doi.org/10.1088/0031-9155/61/4/1677.
[37] Kry SF, Bednarz B, Howell RM, Dauer L, Followill D, Klein E, et al. AAPM TG 158: Measurement and calculation of doses outside the treated volume from external-beam radiation therapy. Med Phys 2017;44:e391–429. https://doi.org/10.1002/mp.12462.
[38] Blanck O, Masi L, Damme M-C, Hildebrandt G, Dunst J, Siebert F-A, et al. Film-based delivery quality assurance for robotic radiosurgery: Commissioning and validation. Phys Medica 2015;31:476–83. https://doi.org/10.1016/j.ejmp.2015.05.001.
[39] Lambrecht M, Melidis C, Sonke J-J, Adebahr S, Boellaard R, Verheij M, et al. Lungtech, a phase II EORTC trial of SBRT for centrally located lung tumours - a clinical physics perspective. Radiat Oncol 2016;11:7. https://doi.org/10.1186/s13014-015-0567-5.
[40] Kron T, Chesson B, Hardcastle N, Crain M, Clements N, Burns M, et al. Credentialing of radiotherapy centres in Australasia for TROG 09.02 (Chisel), a Phase III clinical trial on stereotactic ablative body radiotherapy of early stage lung cancer. Br J Radiol 2018:20170737. https://doi.org/10.1259/bjr.20170737.
[41] Seravalli E, Van Haaren PMA, Van Der Toorn PP, Hurkmans CW. Intercranial stereotactic RT A comprehensive evaluation of treatment accuracy, including end-to-end tests and clinical data, applied to intracranial stereotactic radiotherapy. Radiother Oncol 2015;116:131–8. https://doi.org/10.1016/j.radonc.2015.06.004.

[42] Ishihara T, Yamada K, Harada A, Isogai K, Tonosaki Y, Demizu Y, et al. Hypofraktionierte stereotaktische Strahlentherapie bei Hirnmetastasen eines Lungenkarzinoms: Evaluierung von Indikationen und Prädiktoren der lokalen Kontrolle. Strahlentherapie Und Onkol 2016;192:386–93. https://doi.org/10.1007/s00066-016-0963-2.

[43] Higuchi Y, Serizawa T, Nagano O, Matsuda S, Ono J, Sato M, et al. Three-staged stereotactic radiotherapy without whole brain irradiation for large metastatic brain tumors. Int J Radiat Oncol Biol Phys 2009;74:1543–8. https://doi.org/10.1016/j.ijrobp.2008.10.035.

[44] Wilke L, Andratschke N, Blanck O, Brunner TB, Combs SE, Grosu AL, et al. ICRU report 91 on prescribing, recording, and reporting of stereotactic treatments with small photon beams: Statement from the DEGRO/DGMP working group stereotactic radiotherapy and radiosurgery. Strahlentherapie Und Onkol 2019;195:193–8. https://doi.org/10.1007/s00066-018-1416-x.

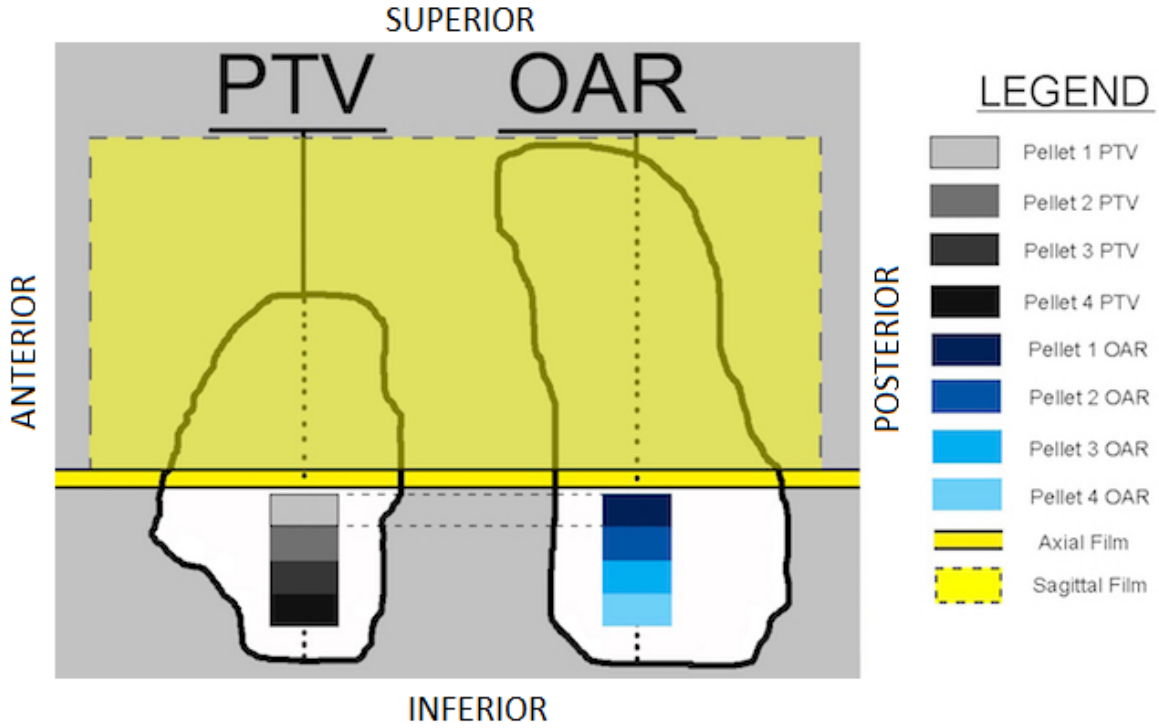


Figure 1: Schematic representation of a sagittal view through the centre of the phantom, showing the positions of the detectors in relation to the target and brainstem structures.

575
576
577
578
579
580
581

Platform No.	Platform	Energy	Technique	Coplanar / Non-coplanar	TPS	Dose Calculation Algorithm	Density Heterog. Correction	Peripheral Prescr. Dose (Gy)	Max Dose (Gy)	Pr Iso. Relative to Dmax (%)
1	VRN TrueBeam STx	10MV FFF	4DCA	NC	BL iPlan	Pencil Beam	Yes	21.0	26.4	80%
2	BL Novalis Tx	6MV	9FF	NC	BL iPlan	Pencil Beam	Yes	17.5	22.7	77%
3	VRN 2100x	6MV	4CCA	NC	BL iPlan	Circular Cone	Yes	18.0	23.6	76%
4	ELK Synergy	6MV	5VMAT	NC	Pinnacle	Convolution	Yes	18.0	22.9	79%
5	ACC Tomotherapy	6MV FFF	Tomo Therapy	C	Tomo Therapy	Non-Voxel Broad Beam	Yes	18.0	20.7	87%
6	BL Novalis Tx	6MV SRS	8FF	NC	Pinnacle	Convolution	Yes	18.0	22.4	80%
7	ELK Agility	6MV	3DCA	NC	Monaco	MC Photon	Yes	18.0	19.9	90%
8	BL Novalis Tx	6MV SRS	4DCA	NC	BL iPlan	Pencil Beam	Yes	18.0	23.1	78%
9	VRN TrueBeam	6MV	4DCA	NC	BL iPlan	Pencil Beam	Yes	18.0	22.8	79%
10	ELK VersaHD	6MV	5VMAT	NC	Monaco	MC Photon	Yes	21.0	26.3	80%
11	VRN TrueBeam	10MV FFF	2VMAT	C	Eclipse	AAA	Yes	20.0	23.3	86%
12	BL Novalis Tx	6MV SRS	4DCA	NC	BL iPlan	Pencil Beam	Yes	18.0	23.2	78%
13	VRN TrueBeam	6MV FFF	1VMAT	C	Eclipse	AAA	Yes	16.0	19.9	80%
14	VRN TrueBeam	10MV FFF	1VMAT	C	Eclipse	AAA	Yes	16.0	20.3	79%
15	ELK BeamMod	6MV	8 FF	NC	Pinnacle	Convolution	Yes	21.0	23.5	89%
16	VRN TrueBeam STx	6MV	5DCA	NC	BL iPlan	Pencil Beam	Yes	18.0	22.7	79%
17	ELK VersaHD	6MV FFF	3VMAT	NC	Monaco	MC Photon	Yes	18.0	34.8	52%
18	VRN TrueBeam	10MV FFF	4VMAT	NC	Eclipse	Acuros	Yes	20.0	30.0	67%
19	ELK BeamMod	6MV	7DCA	NC	Pinnacle	Convolution	Yes	18.0	24.1	75%
20	VRN TrueBeam STx	6MV	5DCA	NC	BL iPlan	Pencil Beam	Yes	21.0	27.1	77%
21	VRN ix2100	6MV	7VMAT	NC	Eclipse	AAA	Yes	21.0	30.5	69%
22	VRN TrueBeam STx	6MV FFF	5DCA	NC	BL iPlan	Pencil Beam	Yes	18.0	22.2	81%
23	ELK GK Perfexion	^{60}Co	17 shots	NC	Gamma Plan	TMR10	No	18.0	40.9	44%
24	ELK GK Perfexion	^{60}Co	19 shots	NC	Gamma Plan	TMR10	No	20.0	40.0	50%
25	ELK GK Icon	^{60}Co	20 shots	NC	Gamma Plan	TMR10	No	18.0	36.7	49%
26	ELK GK Icon	^{60}Co	22 shots	NC	Gamma Plan	TMR10	No	18.0	36.0	50%
27	ELK GK Perfexion	^{60}Co	11 shots	NC	Gamma Plan	TMR10	No	18.0	39.1	46%
28	ELK GK Perfexion	^{60}Co	22 shots	NC	Gamma Plan	TMR10	No	18.0	40.9	44%
29	ELK GK Perfexion	^{60}Co	22 shots	NC	Gamma Plan	TMR10	No	18.0	41.4	43%
30	ACC CK VSI	6MV FFF	138 beams	NC	Multi Plan	Ray Tracing	Yes	21.0	32.3	65%
31	ACC CK VSI	6MV FFF	123 beams	NC	Multi Plan	Ray Tracing	Yes	18.0	25.7	70%

32	ACC CK VSI	6MV FFF	139 beams	NC	Multi Plan	Ray Tracing	Yes	18.0	34.0	53%
33	ACC CK VSI	6MV FFF	109 beams	NC	Multi Plan	Ray Tracing	Yes	20.0	30.8	65%

Table 1: Summary of equipment, techniques and prescription practices of the audit participants. The centres are grouped by platform and in random order, different to the order shown in the results, to avoid identification of individual centres. On-board imaging for positioning the phantom was used by all participants except centres 3, 23, 24, 27, 28 and 29. (VRN=Varian, BL=Brainlab, ELK=Elekta, ACC=Accuray, GK=Gamma Knife, CK=Cyberknife (with cones), DCA=Dynamic Conformal Arcs, FF=Fixed Fields, CCA=Circular Collimator Arcs, VMAT=Volumetric Modulated Arc Therapy)

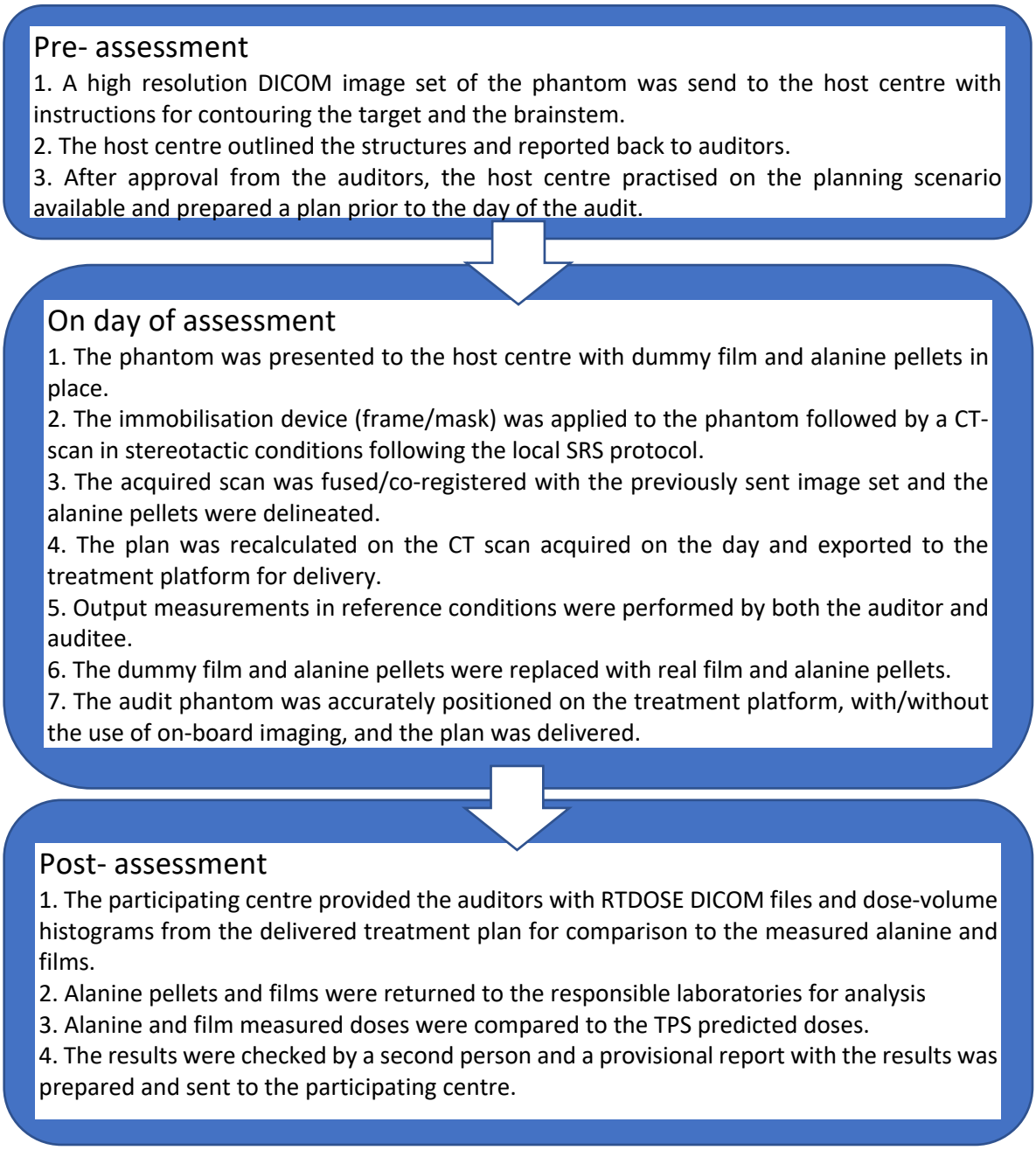


Figure 2: Flowchart diagram showing the main steps of the procedure followed in performing the end-to-end assessments.

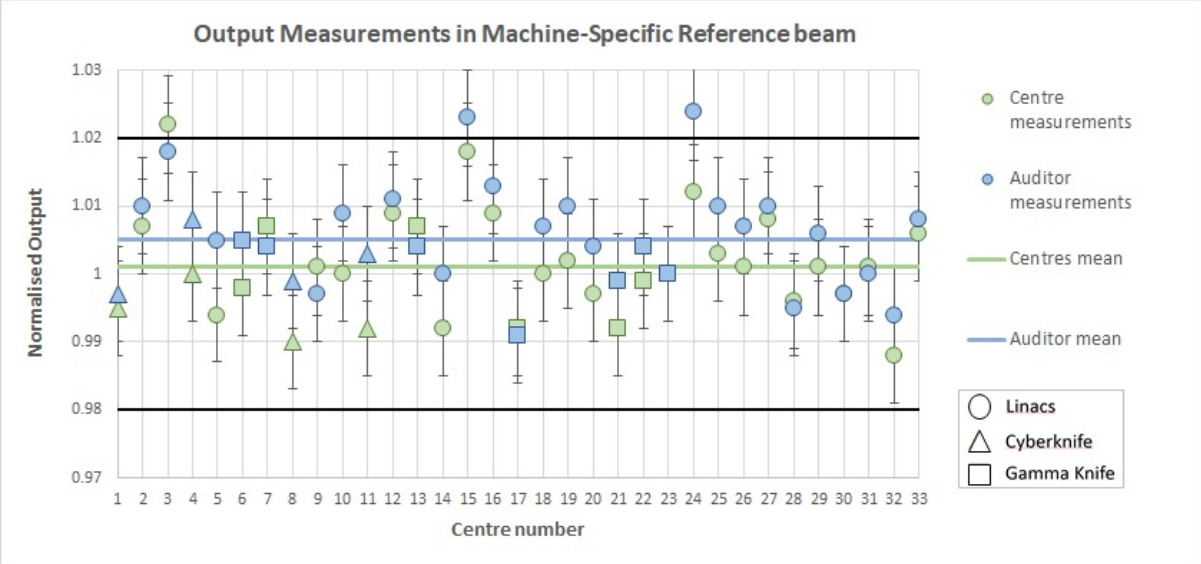


Figure 3: Output measurements in local reference conditions for the 33 platforms that participated in the audit. An uncertainty of ±0.7% (k=1) is indicated by the error bars as a standard protocol based on the calibration certificate of the detector. The “acceptable” tolerances of ±2% are indicated.

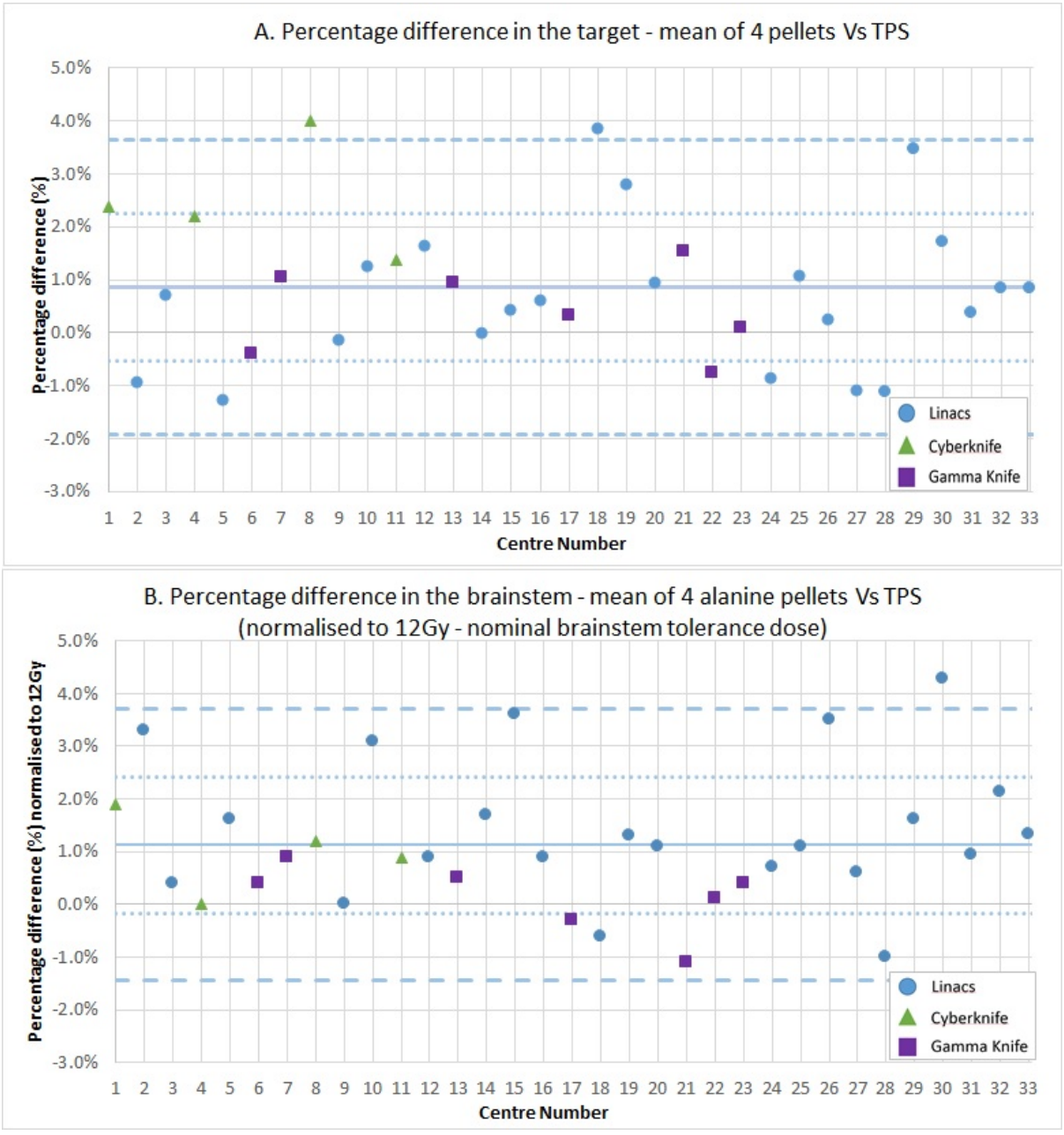


Figure 4: Alanine pellet measurements performed in the target (4A) and organ at risk (4B). Platform groups are indicated in the legends. The mean for all centres is represented by the solid blue line, the dotted lines represent one standard deviation of the mean and the dashed lines represent two standard deviations of the mean.

Short title: Multi-institutional audit of dosimetric delivery in radiosurgery

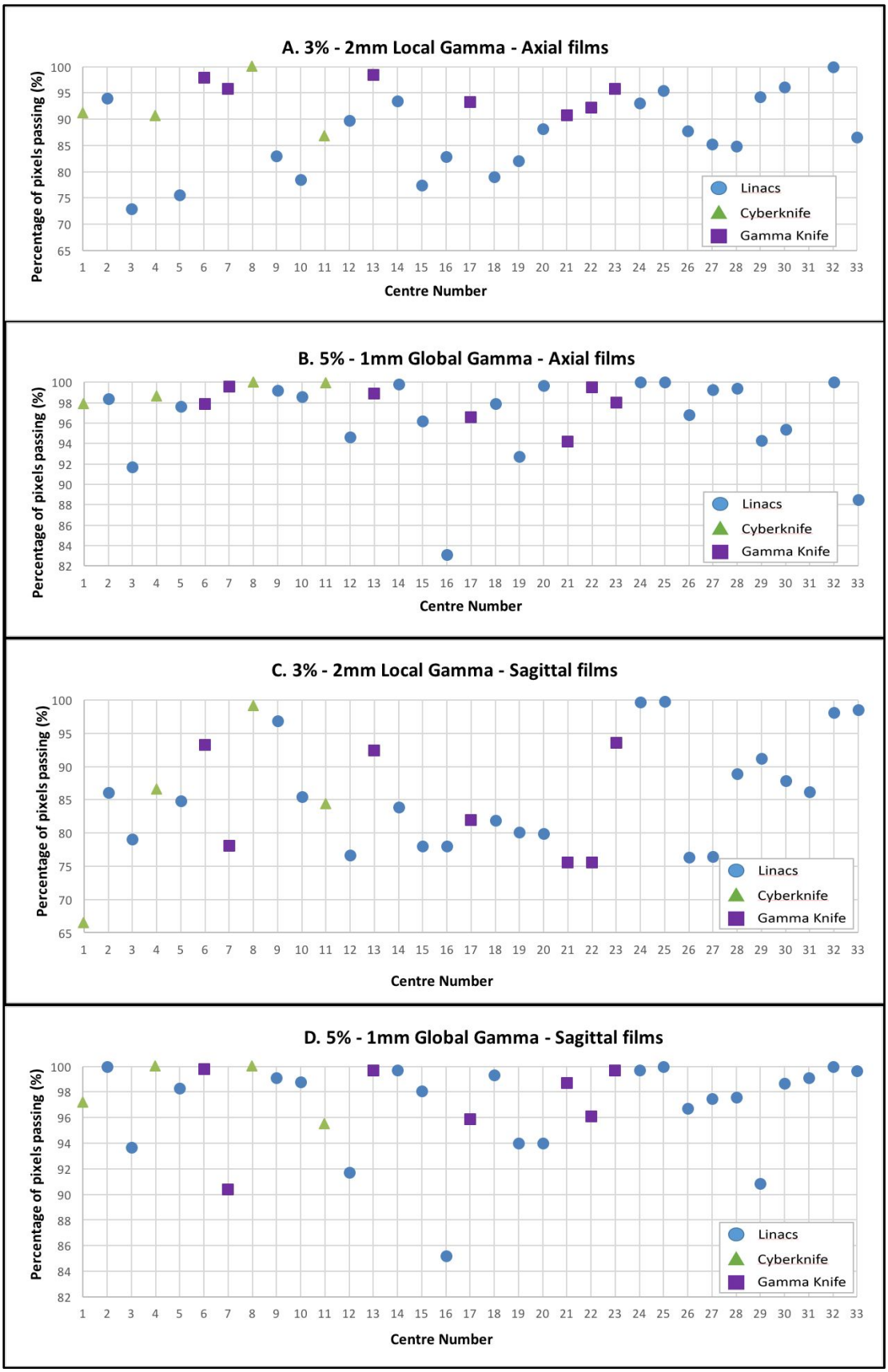
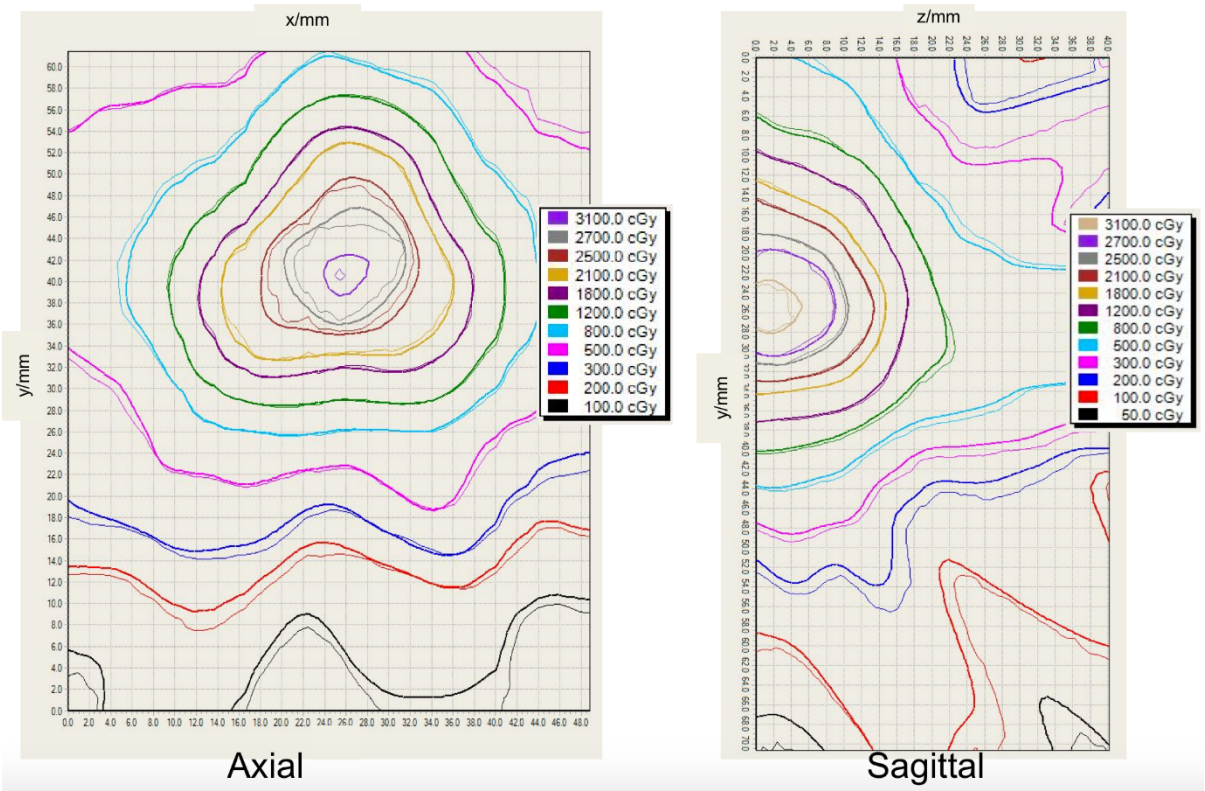
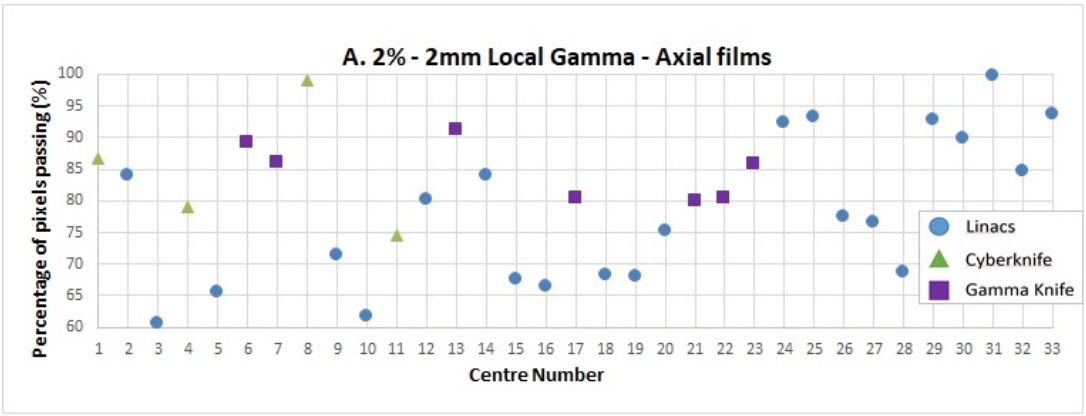
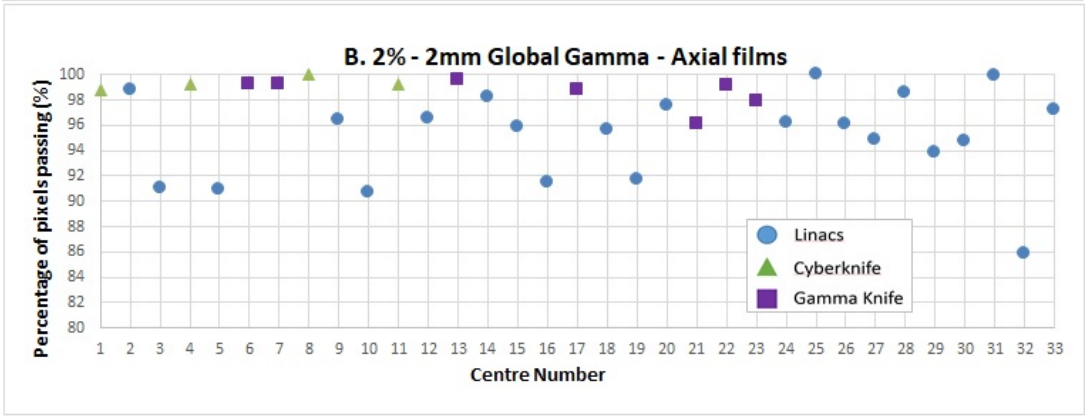


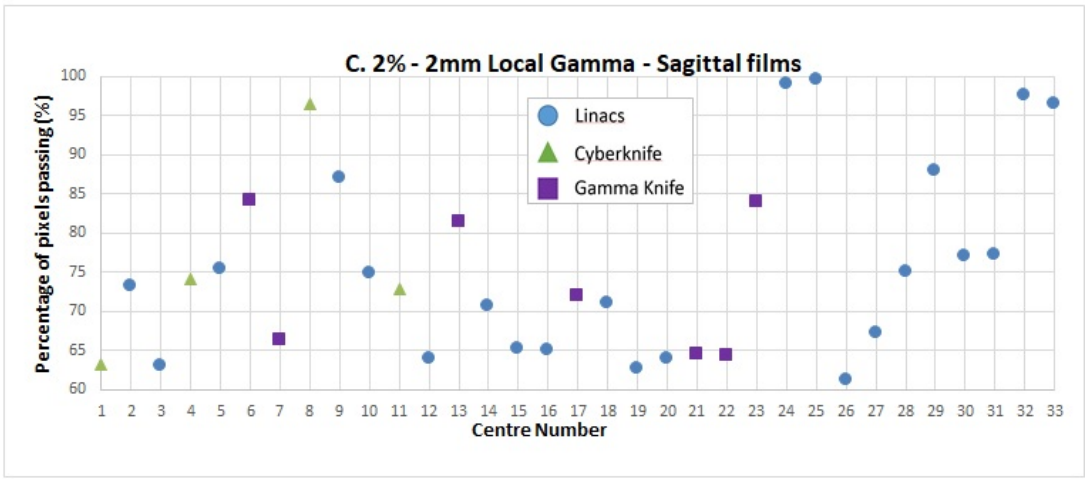
Figure 5: Axial (5A&5B) and Sagittal (5C&5D) film passing rates of the local gamma criterion of 3% - 2mm and the global gamma criterion of 5% - 1mm for the 33 platforms that were assessed.

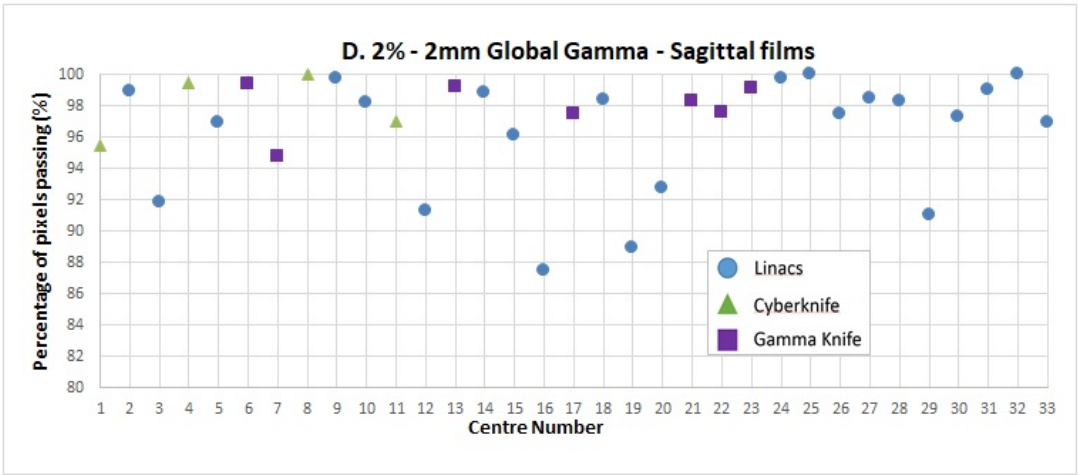


Supplementary Figure 1: Example of dose distribution comparisons between the film-measured doses (thin lines) and the treatment planning system- calculated doses (thick lines) for the axial and the sagittal films used.









683
684 Supplementary Figure 2: Axial (A&B) and Sagittal (C&D) film passing rates of the local gamma and
685 the global gamma criteria of 2% - 2mm the 33 platforms that were assessed.