Title: Development and implementation of an end-to-end test for absolute dose verification of small animal pre-clinical irradiation research platforms

Short title: Dosimetry assessment for pre-clinical irradiations

Conflicts of interest: There are not conflicts of interest to declare from any of the authors

Abstract:

Objectives: Lack of standardization and inaccurate dosimetry assessment in pre-clinical research is hampering the translational opportunities for new radio-therapeutic interventions. The aim of this work was to develop and implement an end-to-end dosimetry test for small animal radiation research platforms to monitor and help improving accuracy of dose delivery and standardization across institutions.

Methods: The test is based on a bespoke zoomorphic heterogeneous mouse and WT1 Petri dish phantoms with alanine as reference detector. Alanine measurements within the mouse phantom were validated with MC simulations at 0.5 mm Cu x-ray reference beam. Energy dependence of alanine in medium x-ray beam qualities was taken into consideration. For the end-to-end test, treatment plans considering tissue heterogeneities were created in Muriplan TPS and delivered to the phantoms at five Institutions using XXX small animal irradiation platforms. Mean calculated dose to the pellets were compared to alanine measured dose.

Results: MC simulations and in phantom alanine measurements in XXX reference beam were in excellent agreement, validating the experimental approach. At one institute, initial measurements showed a larger than 12% difference between calculated and measured dose caused by incorrect input data. The physics data used by the calculation engine were corrected and the TPS was recommissioned. Subsequent end-to-end test measurements showed differences below 5%. With an anterior field, four of the participating

institutes delivered dose to both phantoms within 5%.

Conclusions: An end-to-end dosimetry test was developed and implemented for dose evaluation in preclinical irradiations with small animal irradiation research platforms. The test was capable of detecting treatment planning commissioning errors and it highlighted critical elements in dose calculation. Absolute dosimetry with alanine in relevant pre-clinical irradiation conditions showed reasonable levels of accuracy when compared to TPS calculations. This work is providing an independent and traceable dosimetric validation in pre-clinical research involving small animal irradiations.

Introduction

Advances in pre-clinical radiotherapy research led to the development of precise and sophisticated small animal irradiation platforms([1],[2],[3]) allowing to mimic human radiotherapy conditions on a much smaller scale. These systems may play a critical role for the improvement of our understanding of radiation effects and the development of biologically optimized radiotherapy approaches. However, several concerns have been raised in recent years regarding the quality of published preclinical data which hamper the translational opportunities for a new therapeutic interventions into clinical practice([4]). It has been reported that about half of all preclinical research in the United States is not reproducible([5]). It is well known that radiobiological research is burdened with large systematic uncertainties and variations associated with biological response. Therefore, as stated by Williams *et al.* "given the size of the error in the biological contribution, it is important that the physical errors are minimized"([6]). The lack of reproducibility in pre-clinical radiation studies has been highlighted by Desrosiers *et al.*([7]) back in 2011. Their report, apart from discussing importance of standardization of dosimetry in radiation biology, provides a list of recommendations related to dosimetry content which should be included in any radiobiology and pre-clinical radiation research manuscript. More recently, the ACROP (Advisory Committee in Radiation Oncology Practice) group from ESTRO (European Society for Radiotherapy & Oncology) has commissioned a report for precision small animal radiotherapy research issuing a list of guidelines for reporting studies([8]). However, it will only be possible to evaluate the impact of these guidelines in a few years. A recently published comprehensive review by Draeger *et al.*([9]) demonstrated a deep lack of any reporting of basic physics concepts in pre-clinical radiation research literature, emphasising that a bespoke, robust and standardized way of measuring and reporting dosimetric data and irradiation protocols is still needed to address reproducibility crisis

in radiation biological science and to improve translation of preclinical research into clinical trials. It is clear that existing guidance and recommendations, although informative and comprehensive, are not enough. Researchers should take responsibility for the quality of their data and critically and objectively assess physical quantities that will have an impact on the quality of their findings. Moreover, the requirements on quality assurance procedures and delivered dose verifications in pre-clinical research should aim to reach a level closer to those employed for clinical radiotherapy irradiations. The provision of adequate dosimetric verification tools, including audits, would allow independent assessment of radiation sources used in pre-clinical investigations and provide confidence and standardization in the accuracy of dose evaluation.

Pre-clinical high precision irradiators, such as XXX Small Animal Radiation Research Platform (SARRP)([10]) and X-RAD SmaRT([11]) manufactured by Precision X-Ray, are equipped with cone beam CT (CBCT) imaging panels and treatment planning systems (TPS) making it possible to plan delivery of treatment dose on the acquired CT image. Commissioning of such systems consists of output measurements with a suitable ionization chamber and absolute dosimetry with radiochromic films([12],[13]). Despite their high level of sophistication, there is no independent dose verification process for assessing the accuracy of these irradiators, leaving individual users with the tasks of providing onerous and, often inadequate, quality assurance checks. As a result, based upon procedures existent in clinical radiation therapy([14],[15],[16]), we developed and implemented an end-to-end multi-institutional treatment planning dose verification test. The objective of this study is to describe the procedure, designed to verify dose calculations performed with SARRP's TPS, Muriplan, a system with a superposition-convolution kernel-based calculation algorithm([17]) employing a solid-state alanine dosimeter and phantoms with shapes relevant to pre-clinical irradiations: an anatomically correct mouse phantom, made of tissue equivalent materials([18]) and a WT1 (solid

water) Petri dish-like phantom. The procedure was used to monitor dose deliveries at 5 UK radiobiological centres using SARRPs. The measurements included single and complex field exposures using different size collimators. This work demonstrates how a well-designed dose verification procedure, based on suitable and reliable dosimeters and phantoms, can provide independent and accurate dosimetric assessment of pre-clinical radiation platforms.

Methods and Materials

Five UK institutions, actively involved in preclinical research and equipped with SARRP irradiators, were visited to carry out the end-to-end dosimetry test. For clarity, when presenting the methodology, results and discussions, the following naming convention and abbreviation are used through the paper: Institution 1 (I1), Institution 2 (I2), Institution 3 (I3), Institution 4 (I4) and Institution 5 (I5).

Alanine dosimeters and their energy dependence in SARRP

The XXX alanine measurement service provides alanine dosimeters for absorbed dose to water determination for industrial and radiotherapy applications as well as for research purposes([19]). For this study, the standard size alanine pellets (5 mm in diameter and 2.3 mm in height) were used. The service reports the dose, traceable to XXX primary standard graphite calorimeter for cobalt-60 (^{60}Co) beam quality. Alanine has proven to be a detector with minimal energy dependence (less than 1%) in the range of high energy electrons and photons (for ^{60}Co energy levels and above). Similarly to proton and ion beams, for medium and low energy x-rays, alanine exhibits a much larger energy dependence([20],[21]), which requires correction factors.

Energy dependence correction factors ($r_{Q0,Qx}$), accounting for the lower response of alanine in

medium energy x-ray beams (Q_x), in comparison to ^{60}Co radiation quality (Q_0), were determined in the XXX medium energy reference beams, using the half value layer (HVL) as the beam quality specifier. The experimentally determined $r_{Q0,Qx}$ factors varied from 0.77 to 0.94 for beams with HVL ranging from 0.5 to 4 mm Cu (corresponding to nominal tube potential 135 to 280 kV), respectively. The $r_{Q0,Qx}$ factors, were fitted to the following logarithmic curve with a coefficient of determination, R^2, of 0.9984:

$$r_{Q_0,Q_X} = 0.0829 \times \ln(HVL_{Q_X}) + 0.8266 \qquad (1)$$

To test the validity of alanine energy dependence, given by equation (1), in the users' beam, two sets of measurements were carried out at I1 in SARRP's references conditions([22]), i.e. open field, source-to-isocentre distance (SID) of 35 cm, detector positioned at 2 cm depth and 4 cm of WT1 solid water as underlying backscatter material. Firstly, AAPM's TG-61 code of practice (CoP) absorbed dose to water formalism([23]) was followed to determine the output of the SARRP irradiator (in Gy/s). Measurements were performed with a PTW 30012 ionization chamber, calibrated in terms of air kerma and traceable to the UK primary standard XXX 300 kV free air chamber. Subsequently, a set of 6 alanine pellets were independently irradiated in the same conditions as the ionization chamber measurements. To obtain the value of the average absorbed dose to the alanine volume at SARRP's beam quality, the following equation was used:

$$D_{Q_{SARRP}}^{alanine} = \frac{Dose_{Q_0}^{alanine}}{r_{Q_0,Q_{SARRP}}} \qquad (2)$$

The ($r_{Q_0,Q_{SARRP}}$) energy dependence factor was determined from equation 1 for the I1 SARRP's beam quality of 0.669 mm Cu.

SARRP's output measurements in large field reference conditions, carried out with the ionization chamber and alanine were compared.

Zoomorphic mouse phantom and WT1 (Petri dish) phantom

An anatomically accurate mouse phantom (Figure 1(a)), purposely designed for dosimetry measurements in small animal radiation platforms([18]), was used for this work. The phantom was made of three materials: WT1([24]), Accura Blustone([25]) and LN10([26]), representing soft tissue, cortical bone and lung tissue, respectively(. The phantom included a precisely drilled cavity in the intracranial region to hold a standard XXX alanine pellet. Finally, a removable skull cap (made of WT1 and bone-like material) was built, to be inserted above the alanine pellet.

As part of the validation of the end-to-end test it was decided to include a phantom with a simpler geometry, representing 35 mm diameter Petri dish (Figure 1(b)) made of homogeneous WT1 material with the centre of the analysed pellet positioned at 3.75 mm depth.

Monte Carlo validation of the end-to-end test.

Direct validation of the absolute dose measured by the alanine pellet placed in the mouse phantom was not possible as (i) placing any other type of reference detector (e.g. an ionization chamber) in the animal phantom would significantly alter its scattering properties, which would then require additional corrections for dosimetry assessment, and (ii) the use of the SARRP's dose calculation engine, Muriplan, as reference would defeat the purpose as ultimately the mouse phantom/alanine system is intended to validate the dose delivered, as calculated by such software. As an alternative, we compared the measured and simulated ratios of the doses absorbed by alanine pellets in the phantom and in reference conditions following exposure to a reference radiation beam quality from the XXX 300 kV facility with 7 cm diameter field size. To achieve reference conditions, required by the CoP([23]), the surface of the 30×30×30 cm^3 water phantom was set at a distance of 75 cm from the radiation source. For the small animal phantom, the system with the alanine pellet inserted in the intracranial cavity was set on top of the 30×30×30 cm^3 water phantom (Figure.2). This

approach allowed us to determine the increase in dose experienced by the alanine pellet in the animal phantom relative to the dose delivered to pellet irradiated in reference conditions and to compare it with the simulated value in the same conditions (see Equation 3).

$$\frac{Dose^{measured}_{mouse\ phantom}}{Dose^{measured}_{reference\ conditions}} \text{ vs } \frac{Dose^{MC-simulated}_{mouse\ phantom}}{Dose^{MC-simulated}_{reference\ conditions}} \quad (3)$$

The simulations were performed using the TOPAS (version 3.2) platform, a wrapper which employs the Geant4([27]) simulation toolkit, which have been previously evaluated in low and medium energy x-ray beams([28]). All the simulations performed for this work used the recommended standard physics list as described by TOPAS documentation([29]), standard secondary production cuts and recommended step sizes([30]). Specifically, the secondary particles were produced and transported only when their estimated range was longer than 0.05 mm otherwise their energy was deposited locally. The default maximum step size for particle transport was set to 1 mm. The minimum and maximum particle range were set to 100 eV and 500 MeV, respectively. For the geometry of the simulations, the small animal phantom was imported from the stereolithography (stl) files used for the production of the phantom([18]), with individual files for the skeleton and soft tissue. The relative position of the two anatomical structures and the location of the alanine pellet within the animal skull were determined using images from a CT scan of the manufactured phantom. Contours of the skeleton, alanine pellet and external surface from the CT were then transferred to the projection of the stl files in TOPAS to inform on the relative positioning of the stl anatomical components and to determine the final volume for the alanine pellet.

A previously validated radiation source, accurately representing XXX 0.5 mm Cu beam quality([31]) was modelled in EGSnrc([32]) MC user code BEAMnrc([33]) and the phase space files were used as radiation input for the TOPAS simulations.

Individual simulations were performed using $5 \cdot 10^6$ particle histories and the standard error for the results was computed over the mean of all simulations performed for a specific set up.

End-to-end test workflow

The end-to-end test was designed to verify TPS calculations with different levels of complexity. Firstly, irradiations with the 10×10 mm^2 or with the Ø 8 mm collimator for the institution with bespoke collimators (in both cases covering the full size of the Ø 5 mm alanine pellets), with a single field, in the homogenous WT1 Petri dish phantom and in the mouse phantom, were performed. However, in order to assess the possibility of employing the use of the mouse phantom-alanine assembly for the verification of TPS with smaller field sizes, a set of measurements with the 5×5 mm^2 collimator was also acquired and evaluated. Finally, a set of more pre-clinically relevant irradiations with parallel-opposite, anterior-posterior and arc fields (only in the mouse phantom) were also assessed. In all cases 15 Gy were prescribed to the centre of the pellets. Details of all the radiation exposures are reported in Table 1 in Supplementary Information.

In order to make the process more time efficient, only the segmentation, imaging and irradiation were performed on-site during the visits, whilst calculations of volumes of interest and Muriplan calculated dose were performed a-posteriori. We verified that with the appropriately saved plans and segmentation files, it was possible to contour and to perform calculations of the dose distributions and dose statistics in the ROI a-posteriori without affecting the overall quality and accuracy of the end-to-end test. Figure3 represents the workflow of the procedures developed for the end-to-end test.

Prior to the implementation of the test, the reproducibility of the phantom-alanine assembly was assessed. Fourteen repetitions of the mouse phantom preparation, positioning and imaging as well as pellets' irradiation with a SARRP open field and the same irradiation time, were performed.

The TPS-calculated mean dose to the pellets is affected, both, by the elemental segmentation and the contoured volume of the pellet. The analysis is available in the Supplementary Information.

To evaluate the effects of the segmentation in the calculations, variations on the segmentation threshold (within visually acceptable levels) were performed. Each time, dose distributions and Dose Volume Histograms (DVH) were calculated. The correlation between the segmentation threshold percentage variation with the mean calculated dose was investigated.

Results

Alanine energy dependence in SARRP beam quality

The I1 SARRP's output in reference conditions, measured with the ionization chamber (IC) and electrometer system was 0.0611 Gy/s with an uncertainty of 2.1% (k=1) associated with the determination of absorbed dose to water at 2 cm. (Table III in AAPM TG-61)([23]).

The alanine energy dependence correction factor at SARRP beam quality (HVL of 0.669 mm Cu), calculated from equation 1, was equal to 0.793. Averaged alanine output (for six independent alanine irradiations) measured in SARRP's output reference conditions was 0.0617 Gy/s with an uncertainty of 2.7% (k=1) (associated to the alanine reading process and to the $r_{Q0,Qx}$ experimental setup determination). Both output measurements, carried out with alanine and IC agreed well within the associated uncertainties.

Monte Carlo validation for the end-to-end test

Monte Carlo simulations reported an average value of dose absorbed to the alanine pellet in the intracranial region of the mouse phantom of $(1.232\pm0.030)\times10^{-17}$ Gy/hist. The average dose to medium, scored in the volume of the pellet at reference conditions: depth of 2 cm in a 30×30×30 cm^3 water phantom, along the radiation beam central axis was $(1.090\pm0.054)\times10^{-17}$ Gy/hist. Following equation 3, that leads to a ratio of 1.13±0.06.

Measurements reported an average value of dose over the volume of the alanine pellet for a 0.5 mm Cu HVL of 16.60±0.04 Gy and of 14.60±0.03 Gy in the mouse phantom and in reference conditions, respectively. Following equation 3, the ratio for the measurements was 1.130±0.004.

Evaluation of parameters affecting TPS dose calculations

The mean measured dose to the set of 14 pellets used to assess the repeatability of the phantom-alanine assembly was 35.18 Gy with a standard deviation (SD) of 0.78 Gy. The descriptive statistical analysis showed a confidence level of 0.44 (2σ), which is well below the total uncertainty budget as reported by the XXX alanine service indicating a very good repeatability and stability of the mouse phantom/alanine dosimetry system.

The variation in the segmentation presets, as shown in Figure 4, demonstrate that the presence of voxels inside the pellet's contoured volume, wrongly identified as bone instead of soft tissue material have a significant impact on the calculation of the mean DVH dose. Inadequate bone segmentation threshold, shown in Figure 4 as "Seg_Not_OK" case, gives to rise of the mean calculated dose to the alanine pellet by 13%.

Alanine measurements and comparison with TPS calculations

The HVLs, reported by each institution varied between 0.650 and 0.836, which corresponded to $r_{Q_0,Q_{SARRP}}$ between 0.791 and 0.812, respectively. For I1, a first set of measurements with a single field and the 10×10 mm^2 collimator, showed that the differences between Muriplan dose calculations and alanine measurements were exceeding 12% for both phantoms (Figure 5). As the SARRP output measurements differed from the alanine and ionization chamber measurements by less than 1.1%, additional steps were taken in order to investigate the source of discrepancy between measured and TPS-calculated doses. After advice from the manufacturer, the TPS system

at I1 was recommissioned. Subsequently, a new group of alanine measurements was acquired. The new percentage dose difference between Muriplan dose calculations and alanine measurements was in a good agreement (better than 5%) for both phantoms, indicating that TPS system at I1 was previously not accurately commissioned for the 10×10 mm^2 collimator.

For the rest of the institutions, single field measurements with the 10×10 mm^2 collimator (alternatively the Ø 8 mm collimator) were also performed. Those results are summarized in Figure 6 a) and b) as the percentage (%) dose difference, averaged over the number of irradiated pellets at each institution. Both phantoms are included (for I2, only the mouse phantom was irradiated).

The alanine measurements carried out in the Petri dish (WT1) phantom were, in general, in better agreement with TPS calculations than exposures in the mouse phantom. This could be attributed to larger uncertainties of the Muriplan's calculation engine for more complex tissue geometries([17], [34]). The alanine pellet in the mouse phantom is positioned in the intracranial region, hence the Muriplan needs to model beam propagation through media with varying densities (i.e. bone and soft tissue, representing skull and brain, respectively).

The average dose difference between TPS and measurement among the 38 irradiated alanine pellets in the mouse phantom (all field sizes and beam configurations for all participating centres) was 1.85%. The individual dose points are shown all together in Figure 6 c.

Irradiations in the mouse phantom with a single 5×5 mm^2 field were performed only at two institutes (for details, see Table 2 in Supplementary Information). Only one data point (out of nine) was outside the 5% dose difference margins, however the discrepancy may be associated with the movement of the small animal phantom during the image acquisition process. During CT acquisition, the couch rotates around its central vertical axis and the phantom can move if no

fixation is provided. A larger standard deviation for this test group (in comparison to the irradiations with larger field sizes) could be explained by the challenges present when performing dosimetry with a detector which size is very close to the irradiation field size. The volume averaging effect would have an impact on the measured dose([35]).

Irradiations in the mouse phantom, in more complex beam configurations (i.e. parallel-opposite, anterior-posterior and arc) where performed at three of the participating institutions (details available in Table 2 in Supplementary Information). Three data points (out of seven) were above 5% dose difference limit. These differences can be attributed to a number of reasons including (i) volume averaging effect of alanine pellets when exposed to small field sizes, (ii) position of the couch within the beam during treatment or (iii) inaccuracies in commissioning of the TPS.

Discussion and conclusions

We developed an end-to-end treatment planning dose verification test for small animal pre-clinical research platform employing an alanine dosimeter and phantoms made of tissue equivalent materials. The test was implemented in a multi-institutional comparison. We presented the results of the differences between the dose calculated by the TPS and dose measured. Apart from one institute, the differences between TPS-calculated and measured doses in both phantoms are below 5%. These discrepancies can be attributed to several reasons. First of all, alanine dosimeters require energy response correction for medium energy x-ray beam. The $r_{Q_0,Q_{SARRP}}$ factor is dependent on the accuracy of beam quality (HVL) determination and for different SARRP machines, used in this investigation, that factor varied by up to 2.6% depending on the irradiator used. HVL determination in the small animal irradiators is certainty not trivial due to space constraints. Moreover, the spectra of the SARRP irradiators can differ significantly from XXX reference beam qualities, hence the

$r_{Q_0,Q_{SARRP}}$ factor calculated from equation 3 will have an additional associated uncertainty. However, previous work carried out at XXX (see Supplementary Information) has shown that for a given HVL, even significant changes in the x-ray spectra, would have very small impact on alanine response (~1% variation). Moreover, there are a number of limitations of Muriplan that could impact the accuracy of the calculated dose([17]), namely (i) the higher level of complexity needed from the calculation algorithms to handle the strong atomic number and energy dependence on the photon interaction cross section, typical for the kilovoltage energy range([36]), (ii) the need for a selection of appropriate tissue segmentation thresholds([37]) and (iii) last, but not less important, the uncertainties related to the data used for the commissioning of very small fields.

It is difficult and also out of the scope of this work to interpret the results acquired at each institute independently as the only common factor for all the data presented in this work are the alanine in-phantom measurements, which were acquired, processed and analysed in the same way. Although the authors had no influence on the TPS commissioning at any of the participating centres, this work demonstrated that the presented end-to-end test can detect errors in the TPS commissioning data and highlight not-correct procedures. By discussing the results of the independent dose verification measurements with the individual institutions and with the manufacturer, it is possible to improve the accuracy of the dose delivered in preclinical irradiations (see Figure 5) and to optimize irradiation procedures. Due to the level of sophistication and complexity of treatment planning systems, before commencing any new pre-clinical study, additional verification of the calculation engines should be required. For the smaller field sizes (below 5 mm), the dosimetry system based on 5 mm in diameter alanine pellet will not provide accurate measurements due to volume averaging effect. Application of smaller size alanine pellets could provide solution to this limitation, however such system would require much larger dose to be delivered (100 Gy and

above) to allow for accurate readout.

This work demonstrates how a well-designed dose verification procedure can provide independent and accurate assessment of pre-clinical irradiation platforms resulting in a tool for implementing the recent recommendations on pre-clinical dosimetry([8]). Moreover, the results provide a degree of standardization, which increases confidence for the comparison of the radiobiological studies performed at the 5 participating UK centres. This dosimetry system could also be used for dose verification in irradiators with different calculation algorithms as well as in simple x-rays cabinets, where a TPS is not available([38]). Moreover, this system could be translated to different radiation types given that tissue equivalence of the phantom and energy dependence of alanine is properly evaluated. In the future, more effort should be made to incorporate additional 2D assessment, e.g. by employing radiochromic films or 2D arrays. This will, however, require additional input from vendors of the pre-clinical IGRT systems to allow for easy export of 2D dose distributions.

References

1. Wong J, Armour E, Kazanzides P, et al. High-Resolution, Small Animal Radiation Research Platform With X-Ray Tomographic Guidance Capabilities. *Int J Radiat Oncol Biol Phys*. 2008;71(5):1591-1599. doi:10.1016/j.ijrobp.2008.04.025
2. Clarkson R, Lindsay PE, Ansell S, et al. Characterization of image quality and image-guidance performance of a preclinical microirradiator. *Med Phys*. 2011;38(2):845-856. doi:10.1118/1.3533947
3. Song KH, Pidikiti R, Stojadinovic S, et al. An x-ray image guidance system for small animal stereotactic irradiation. *Phys Med Biol*. 2010;55(23):7345-7362. doi:10.1088/0031-9155/55/23/011

4. Stone HB, Bernhard EJ, Coleman N, et al. Preclinical Data on Efficacy of 10 Drug-Radiation Combinations: Evaluations, Concerns, and Recommendations 1. *Transl Oncol*. 2016;9:46-56. doi:10.1016/j.tranon.2016.01.002

5. Irreproducible life sciences research in U.S. costs $28 billion | Science News. https://www.sciencenews.org/article/irreproducible-life-sciences-research-us-costs-28-billion. Accessed October 9, 2019.

6. Williams JP, Brown SL, Georges GE, et al. Animal Models for Medical Countermeasures to Radiation Exposure. *Radiat Res*. 2010;173(4):557-578. doi:10.1667/rr1880.1

7. Desrosiers M, DeWerd L, Deye J, et al. The Importance of Dosimetry Standardization in Radiobiology. *J Res Natl Inst Stand Technol*. 2013;118:403. doi:10.6028/jres.118.021

8. Verhaegen F, Dubois L, Gianolini S, et al. ESTRO ACROP: Technology for precision small animal radiotherapy research: Optimal use and challenges. *Radiother Oncol*. 2018;126(3):471-478. doi:10.1016/j.radonc.2017.11.016

9. Draeger E, Sawant A, Johnstone C, et al. A Dose of Reality: How 20 Years of Incomplete Physics and Dosimetry Reporting in Radiobiology Studies May Have Contributed to the Reproducibility Crisis. *Int J Radiat Oncol*. 2019. doi:10.1016/j.ijrobp.2019.06.2545

10. SARRP | Xstrahl Medical & Life Sciences. https://xstrahl.com/life-science-systems/small-animal-radiation-research-platform/. Accessed October 9, 2019.

11. X-RAD SmART - Precision X-Ray | accela. https://www.accela.eu/precision-x-ray/x-rad-smart. Accessed October 9, 2019.

12. Feddersen T V., Rowshanfarzad P, Abel TN, Ebert MA. Commissioning and performance characteristics of a pre-clinical image-guided radiotherapy system. *Australas Phys Eng Sci Med*. 2019;42(2):541-551. doi:10.1007/s13246-019-00755-4

13. Wang YF, Lin SC, Na YH, Black PJ, Wuu CS. Dosimetric verification and commissioning for a small animal image-guided irradiator. *Phys Med Biol*. 2018;63(14). doi:10.1088/1361-6560/aacdcd

14. Wesolowska P, Georg D, Lechner W, et al. Testing the methodology for a dosimetric end-to-end audit of IMRT/VMAT: results of IAEA multicentre and national studies. *Acta Oncol (Madr)*. 2019;0(0):1-9. doi:10.1080/0284186x.2019.1648859

15. Seravalli E, Van Haaren PMA, Van Der Toorn PP, Hurkmans CW. A comprehensive evaluation of treatment accuracy, including end-to-end tests and clinical data, applied to intracranial stereotactic radiotherapy. *Radiother Oncol*. 2015;116(1):131-138. doi:10.1016/j.radonc.2015.06.004

16. Poder J, Brown R, Porter H, Gupta R, Ralston A. Development of a dedicated phantom for multi-target single-isocentre stereotactic radiosurgery end to end testing. *J Appl Clin Med Phys*. 2018;19(6):99-108. doi:10.1002/acm2.12452

17. Reinhart AM, Fast MF, Ziegenhein P, Simeonnill, Oelfke U. A kernel-based dose calculation algorithm for kV photon beams with explicit handling of energy and material dependencies. *Br J Radiol*. 2017;90(1069). doi:10.1259/bjr.20160426

18. Soultanidis G, Subiel A, Renard I, et al. Development of an anatomically correct mouse phantom for dosimetry measurement in small animal radiotherapy research. *Phys Med Biol*. 2019;64(12). doi:10.1088/1361-6560/ab215b

19. XXX

20. Waldeland E, Hole EO, Sagstuen E, Malinen E. The energy dependence of lithium formate and alanine EPR dosimeters for medium energy x rays. *Med Phys*. 2010;37(7):3569-3575. doi:10.1118/1.3432567

21. Anton M, Büermann L. Relative response of the alanine dosimeter to medium energy x-rays. *Phys Med Biol*. 2015;60(15):6113-6129. doi:10.1088/0031-9155/60/15/6113

22. Tryggestad E (Xstrahl), Burton D (Xstrahl), Lu B (Xstrahl). *Dosimetric Commissioning of the SARRP. V10.*; 2016.

23. Ma CM, Coffey CW, DeWerd LA, et al. AAPM protocol for 40-300 kV x-ray beam dosimetry in radiotherapy and radiobiology. *Med Phys*. 2001;28(6):868-893. doi:10.1118/1.1374247

24. White DR, Booz J, Griffith R V., Spokas JJ, Wilson IJ. *ICRU Report 44. Tissue Substitutes in Radiation Dosimetry and Measurements*. Vol os23. SAGE Publications; 1989. doi:10.1093/jicru/os23.1.1

25. 3D Systems Corporation. Tech specs | 3D Systems. https://www.3dsystems.com/materials/accura-bluestone/tech-specs. Published 2015. Accessed August 29, 2019.

26. Leeds Test Objects Ltd. Tissue Equivalent Materials The Original Tissue Equivalents. https://www.leedstestobjects.com. Published 2014. Accessed August 29, 2019.

27. Agostinelli S, Allison J, Amako K, et al. GEANT4 - A simulation toolkit. *Nucl Instruments Methods Phys Res Sect A Accel Spectrometers, Detect Assoc Equip*. 2003;506(3):250-303. doi:10.1016/S0168-9002(03)01368-8

28. Hewson EA, Butson MJ, Hill R. Evaluating TOPAS for the calculation of backscatter factors for low energy x-ray beams. *Phys Med Biol*. 2018;63(19). doi:10.1088/1361-6560/aadf28

29. Perl J, Shin J, Schümann J, Faddegon B, Paganetti H. TOPAS: An innovative proton Monte Carlo platform for research and clinical applications. *Med Phys*. 2012;39(11):6818-

6837. doi:10.1118/1.4758060

30. Jarlskog CZ, Paganetti H. Physics settings for using the Geant4 toolkit in proton therapy. *IEEE Trans Nucl Sci*. 2008;55(3):1018-1025. doi:10.1109/TNS.2008.922816

31. Bass GA, Duane S, Kelly M, et al. *The NPL Air Kerma Primary Standard Free-Air Chamber for Medium Energy X-Rays: Summary of Factors Incorporating ICRU Report 90 Recommendations*. Teddington; 2019. http://eprintspublications.npl.co.uk/8504/#.XcBa5BtVeO4.mendeley. Accessed November 4, 2019.

32. Kawrakow I, Mainegra-Hing E, Rogers DWO, Tessier F, Walters BRB. The EGSnrc Code System: Monte Carlo Simulation of Electron and Photon Transport, NRCC Report PIRS-701. *Natl Res Counc Canada*. 2019:2001-2019. https://nrc-cnrc.github.io/EGSnrc/doc/pirs701-egsnrc.pdf.

33. Rogers DWO, Walters B, Kawrakow I. *BEAMnrc Users Manual*.; 2005.

34. Vaniqui A, Walters BR, Fonseca GP, Verhaegen F. Dose to water versus dose to medium from cavity theory applied to small animal irradiation with kilovolt x-rays. *Phys Med Biol*. 2019;64(16):165001. doi:10.1088/1361-6560/ab2db1

35. Wuerfel JU. Dose measurements in small fields. *Med Phys Int J*. 2013;1(1):81-90. http://mpijournal.org/pdf/2013-01/MPI-2013-01-p081.pdf.

36. Pawlowski JM. Development of a new model-based radiation dose calculation algorithm for kilovoltage energy x-rays. *Vanderbilt Univ Electron Theses Diss*. 2013. https://etd.library.vanderbilt.edu/available/etd-06012013-141128/unrestricted/Pawlowski.pdf.

37. Biglin ER, Price GJ, Chadwick AL, Aitkenhead AH, Williams KJ, Kirkby KJ. Preclinical

dosimetry: Exploring the use of small animal phantoms. *Radiat Oncol*. 2019;14(1):1-10. doi:10.1186/s13014-019-1343-8

38. Kuess P, Bozsaky E, Hopfgartner J, Seifritz G, Dörr W, Georg D. Dosimetric challenges of small animal irradiation with a commercial X-ray unit. *Z Med Phys*. 2014;24(4):363-372. doi:10.1016/j.zemedi.2014.08.005

Figure Captions

Figure 1. Mouse phantom and CBCT images (a) and WT1 Petri dish phantom with associated CBCT (b). Alanine pellets have soft tissue equivalent properties; hence they have similar CT number as WT1 material.

Figure 2. Set-up for measurements and Monte Carlo simulations of the ratio of absorbed dose in alanine pellets inside the mouse phantom (bottom) versus references conditions (top).

Figure 3. Workflow of the end-to-end test.

Figure 4. Examples of four different SARRP segmentation files capturing alanine pellet (circular region below the table) with four different presets ("Seg_OK", "Seg_OK_a", "Seg_OK_b" and "Seg_Not_OK") for the same CBCT acquisition, demonstrating the effect of an inadequate segmentation on the calculated mean dose to the alanine pellets.

Figure 5. Percentage dose difference results (TPS versus alanine measurements) at I1, before and after the TPS recommissioning.

Figure 6. Average percentage dose difference between the TPS calculation and alanine measurements for the mouse (a) and WT1 Petri-dish (b) phantoms for anterior fields only. The error bars represent the standard deviation of the mean. The percentage dose difference for all exposures in the mouse phantom (field sizes > 5mm × 5mm and all beam configurations) and all participating institutes (c).