



Brief Report

Preclinical Tumor Growth Delay Is More Reliable from Imaging-Based Rather than Manual Caliper Volume Measurements

Ifeanyichukwu Ogobuiro , Benjamin Spieler and Ivaylo B. Mihaylov *

Department of Radiation Oncology, Leonard M. Miller School of Medicine, University of Miami, 1475 NW 12th Ave., Suite 1500, Miami, FL 33136, USA; ico8@miami.edu (I.O.); bspieler@miami.edu (B.S.)

* Correspondence: i.mihaylov@med.miami.edu

Abstract

Background/Objectives: Tumor growth delay is frequently used in preclinical experiments evaluating oncologic interventions. While treatment response in humans is based on imaging criteria for obvious reasons, manual caliper measurement of subcutaneous tumors is standard in animal studies. In a murine tumor model treated with immunotherapy (ImT) and radiotherapy (RT), the reliability of caliper measurements was tested by comparing normalized tumor growth delay (NTGD) rates derived from caliper- and image-based volumetrics. **Methods:** A 4T1 breast syngeneic murine model was used, in which thirty animals were inoculated in the right inguinal mammary fat pad and the right axilla. One RT fraction of 8 Gy was delivered to the right inguinal tumor on day 11 post-implant, and intraperitoneal ImT (PD-1 checkpoint inhibitor) injections were administered on days 11, 12, and 14. Each animal underwent three MRI scans (days 10, 17, and 20). Caliper measurements were also performed by two independent observers on the same days. The measurements were averaged and used to estimate ellipsoid tumor volumes. The acquired MRIs were used for image segmentation and volume estimation. Tumor volumes (days 17 and 20) were normalized against the baseline pre-treatment tumor volume (day 10). NTGD rates derived from hand- and image-based volumetrics were compared to assess the reliability of caliper vs. MRI estimation. **Results:** Caliper volumes between the two observers correlated at 0.799 (Pearson, $p < 0.001$). The averaged caliper volumes correlated with MRI volumes at 0.897 (Pearson, $p < 0.001$). Absolute volume differences between caliper and MRI increased with tumor growth. NTGD-derived rates showed no correlation, with only 15% of NTGD caliper rates falling within 10% of the MRI rates. **Conclusions:** NTGD rate based on caliper volumes is a suitable measure of treatment response in preclinical studies. In the experiment described herein, caliper-derived NTGD rates did not correlate with MRI ground truth. These findings suggest that more accurate tumor volumetrics, derived from stored and verifiable medical imaging sources, should be used in preclinical assessment of oncologic interventions instead of standard caliper estimates.

Keywords: murine; growth; measurement; MRI; caliper



Academic Editor: Bing Guo

Received: 4 September 2025

Revised: 9 October 2025

Accepted: 11 October 2025

Published: 14 October 2025

Citation: Ogobuiro, I.; Spieler, B.; Mihaylov, I.B. Preclinical Tumor Growth Delay Is More Reliable from Imaging-Based Rather than Manual Caliper Volume Measurements. *Biomedicines* **2025**, *13*, 2503.

<https://doi.org/10.3390/biomedicines13102503>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The use of murine subcutaneous tumor models for evaluating the efficacy of novel therapeutics by measuring tumor growth delay is a common approach [1]. The subcutaneous tumor model allows for relatively easy longitudinal monitoring of tumor growth

delay. Calipers are commonly utilized to measure tumor volume because they are fast, inexpensive, and do not require anesthesia [2]. However, the use of calipers to longitudinally measure tumor volume has been shown to be error-prone for several reasons [3]. First, caliper measurements are commonly collected along the longest two dimensions of the tumor in the x- and y-planes, with the assumption that the z-axis is the same as the shortest dimension, creating an ellipsoidal shape, which is often inaccurate [2]. Second, measurement from the epidermis, adipose tissue, and fur (if present) introduces error and variability in tumor volume measurements [4]. These relatively simple reasons, but with far-reaching consequences, underscore the need to identify improved methods for measuring tumor volumes accurately and reproducibly in murine subcutaneous tumor models.

Several methods have emerged as alternatives to calipers for measuring subcutaneous tumor volume. Casting of the tumor region has been used to estimate tumor volume but has wide variability, and requires specific locations, such as the thoracic region, which provides a rigid base to capture the tumor volume fully in the casting [5]. Next is bioluminescence imaging (BLI), which is used to measure relative tumor volume. BLI, though relatively rapid with high throughput, only provides a two-dimensional planar format and does not capture absolute tumor volume. Also, BLI requires the tumor cell line to express luciferase, which limits its utility. Alternatively, computed tomography (CT) has been used to measure tumor volume and is more accurate than calipers [6,7]. However, CT is limited by inherently poor soft-tissue contrast when exogenous contrast materials are not used, and by its relatively high cost. Furthermore, CT imaging can introduce confounding factors during longitudinal tumor volume measurement via ionizing radiation. Lastly, ultrasound (US) imaging has been utilized. US has several advantages: it is inexpensive, noninvasive, provides good soft-tissue contrast without exogenous contrast, has high throughput with reliable tumor-volume measurement, and does not use ionizing radiation [8,9]. However, US imaging is less effective for small tumors, is susceptible to artifacts, and requires a steep learning curve for proficient use [10]. These limitations highlight the advantages of using MRI, which has been shown to capture small tumors with high accuracy and intrinsic contrast while maintaining high throughput and no ionizing radiation, making it the ground truth [11–13]. However, most preclinical subcutaneous tumor-volume measurements default to the use of calipers, given the high cost of MRI.

With MRI widely recognized as the ground truth for measuring tumor growth delay, the main objective of this work is to evaluate how well caliper measurements reflect MRI-based tumor-volume measurements.

2. Materials and Methods

The study was performed in a syngeneic, orthotopic 4T1 murine breast tumor model. The breast cancer cells were implanted in the inguinal mammary fat pad and right axilla. Subsequently, the animals were exposed to a single RT fraction of 8 Gy followed by intraperitoneal ImT injections on three consecutive days. Tumor size was then longitudinally measured by both caliper and MRI. Afterward, normalized tumor volumes were compared to assess the accuracy of caliper use.

Cell lines: 4T1, a mouse-derived breast cancer cell line, which mimics human stage IV breast cancer in BALB/c mice, was purchased from ATCC (#CRL-2539). The cells were cultured in RPMI-1640 medium (Gibco; Thermo Fisher Scientific, Inc., Waltham, MA, USA) supplemented with 10% fetal bovine serum (Gibco; Thermo Fisher Scientific, Inc., Waltham, MA, USA) and 1% penicillin–streptomycin (HyClone; GE Healthcare Life Sciences, Logan, UT, USA).

Murine model: Female BALB/c mice ($n = 30$; 8 weeks old; 18–20 g) were purchased from Jackson Laboratory (Bar Harbor, ME, USA). The animals were housed at $22 \pm 5^\circ\text{C}$ on a 12 h light/dark cycle and fed rodent chow and water ad libitum at the University of Miami research animal facility. All experiments were approved by the Institutional Animal Care and Use Committee (IACUC), protocol 17-214-ad02 EDR.

Experimental design: The experimental timeline is presented in Figure 1. First, 5×10^6 4T1 breast cancer cells in a 100 μL mixture were injected orthotopically into the fifth mammary (primary tumor) pad on the right side and subcutaneously on the upper shoulder (abscopal tumor) on the right side of BALB/c mice. A single 8 Gy fraction of radiation therapy was delivered to the right inguinal mammary fat-pad tumor on day 11 post-implant, and intraperitoneal injections of anti-PD-1/CTLA-4 at 20 mg/kg were administered on days 11, 12, and 14. Each animal underwent caliper- and MRI-based tumor-volume measurements on days 10, 17, and 20. Tumor volumes measured on days 17 and 20 were normalized to the baseline pretreatment tumor volume (day 10). Normalized tumor growth delay (NTGD) rates derived from caliper- and MRI-based volumetrics were compared to assess the reliability of calipers versus MRI, as previously shown [14,15]. Of note, NTGD is defined as the fractional change from baseline: $\text{NTGD} = V_t/V_0$, where V_0 = tumor volume at baseline (day 10) and V_t = tumor volume at follow-up (day 17 or 20). Thus, $\text{NTGD} = 1$ means no change; $\text{NTGD} > 1$ indicates growth; and $\text{NTGD} < 1$ indicates shrinkage.

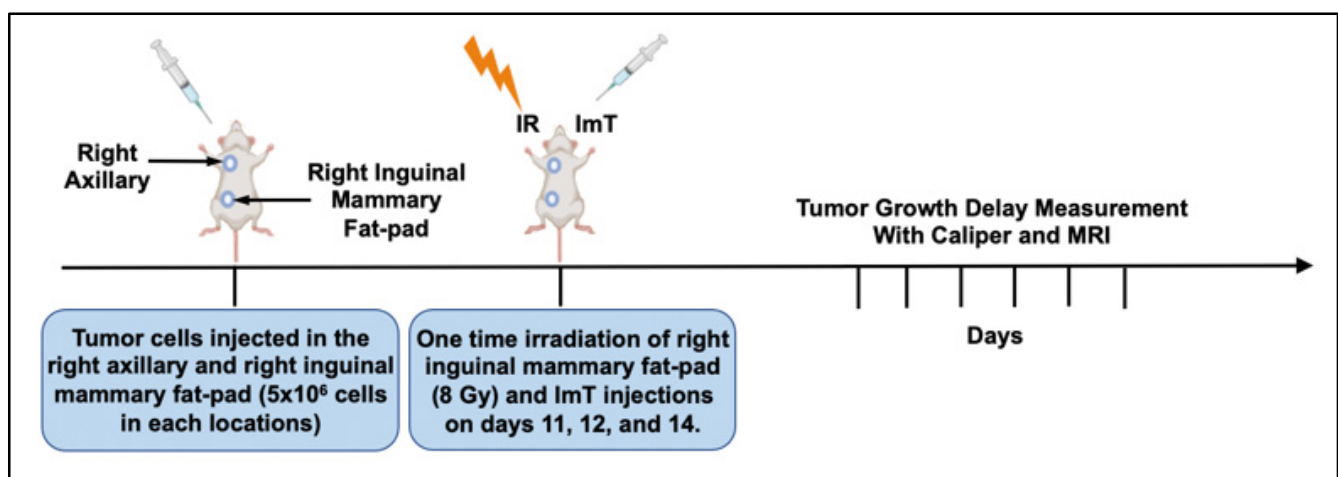


Figure 1. Experimental design of in-vivo experiments aimed to evaluated accuracy of tumor growth delay volume measurement of caliper compared to MRI after irradiation and immunotherapy.

Caliper measurements of subcutaneous tumors: A digital caliper was used to measure each subcutaneous tumor while the mice were conscious. The two longest perpendicular axes in the x- and y-planes of each tumor were measured by two independent observers. The y-axis was assumed to be the same as depth, and tumor volume was calculated using Equation (1), a standard practice.

$$V_{st} = xy^2/2 \quad (1)$$

MRI imaging and data analysis: Animals were subjected to magnetic resonance imaging (MRI) on days 10, 17, and 20 after tumor inoculation. Mice were kept under anesthesia using ketamine administered intraperitoneally at 90 mg/kg, pre-mixed with xylazine (#59399-110-20; AnaSed Injection, Akorn, Lake Forest, IL, USA) at 4.5 mg/kg. Throughout imaging, body temperature was maintained using HotSnapz reusable heating pads. At the end of the procedure, atipamezole hydrochloride (Zoetis) was injected subcutaneously at 7.5 mg/kg to aid recovery from anesthesia. After injection, the animals were placed back in their

cages for recovery. For MR imaging, a T1 sequence with a voxel size of $0.5 \times 0.5 \times 0.5 \text{ mm}^3$ was used on a 3T Siemens TrioTim scanner (Siemens Medical Solutions, Malvern, PA, USA).

Statistical analysis: Non-normalized and normalized caliper- and MRI-based tumor-volume measurements were compared, and differences were considered statistically significant at $p < 0.05$. Pairwise linear fits were plotted, with Pearson R^2 computed in Excel. Statistical analyses were performed in Prism version 9.1.2 (GraphPad Software Inc., San Diego, CA, USA).

3. Results

The technician variability was relatively small for both the axillary and the mammary fat-pad tumor volumes. For the right axillary site, the average caliper measurement on day 10 for technician A and technician B was 415.93 mm^3 ($284.9\text{--}487.70 \text{ mm}^3$) and 347.54 mm^3 ($269.50\text{--}400.5 \text{ mm}^3$), respectively ($p = 0.08$), and the average caliper measurement at day 2 was 1008.88 mm^3 ($486\text{--}1206.83 \text{ mm}^3$) and 906.75 mm^3 ($630.2\text{--}1177.68 \text{ mm}^3$), respectively ($p = 0.50$) (Figure 2A,B). For the right mammary fat pad, the average caliper measurement at day 10 for technician A and technician B was 330.75 mm^3 ($203.15\text{--}425.30 \text{ mm}^3$) and 305.77 mm^3 ($210.8\text{--}401.45 \text{ mm}^3$), respectively ($p = 0.48$), and the average caliper measurement at day 20 was 879.92 mm^3 ($497.13\text{--}1044 \text{ mm}^3$) and 781.47 mm^3 ($599.75\text{--}928.23 \text{ mm}^3$), respectively ($p = 0.43$) (Figure 2C,D).

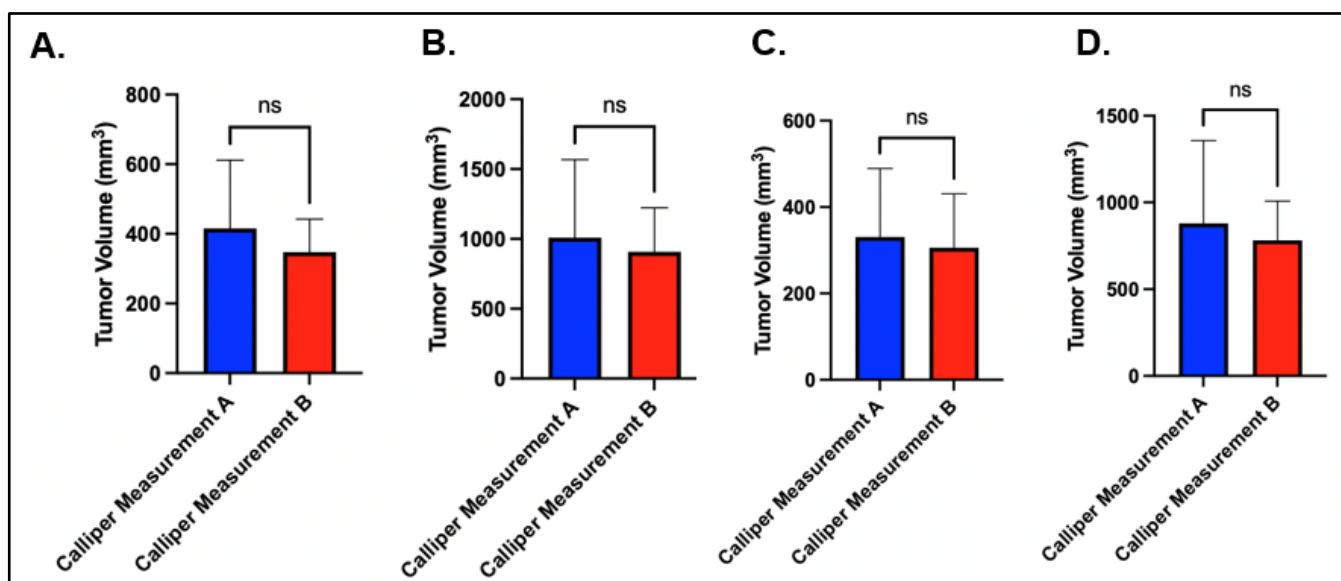


Figure 2. Caliper-based tumor measurements comparing technicians A and B for: (A) day 10 (treatment onset) right axillary; (B) day 20 right axillary; (C) day 10 right mammary fat pad; (D) day 20 right mammary fat pad.

Correlation and Bland–Altman analyses were performed to compare tumor growth-delay volume calculated from caliper measurements (using the ellipsoidal formula) versus estimation by MRI. Analysis of 30 paired caliper and MRI measurements, with interquartile ranges of $183.78\text{--}664.35 \text{ mm}^3$ and $274.18\text{--}842.09 \text{ mm}^3$, respectively, yielded a correlation coefficient of 0.897 (Pearson, $p < 0.001$) (Figure 3A).

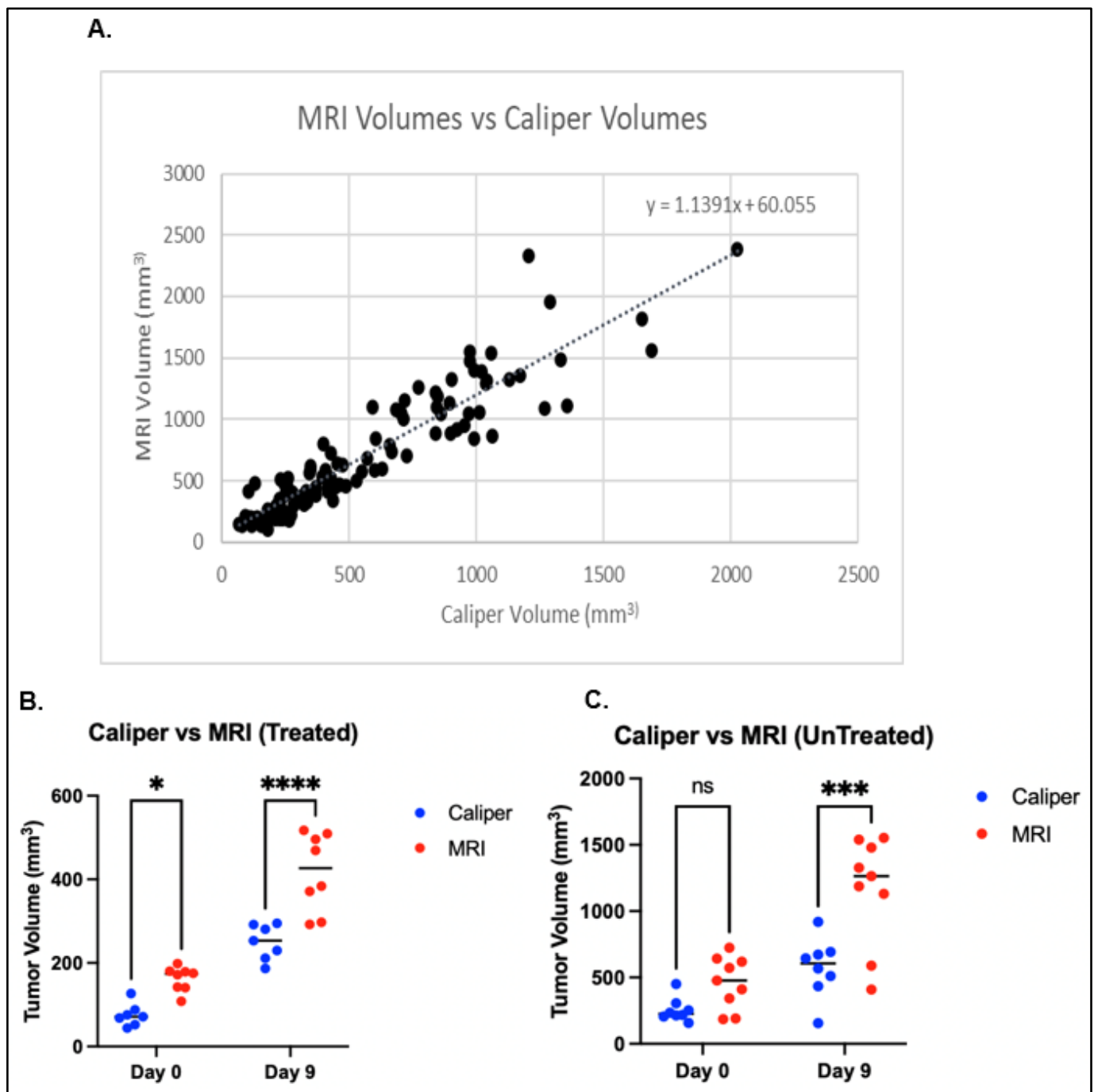


Figure 3. MRI volumes versus caliper-based volumes. (A) Tumor volume measured by MRI versus caliper. (B) Tumor volume measured by MRI versus caliper in treated (radiation and immunotherapy) tumors. (C) Tumor volume measured by MRI versus caliper in untreated (no radiation or immunotherapy) groups. ns means no statistical significance, while asterisk indicates statistical significance.

Although there was a positive correlation in tumor volume, there were significant differences in the absolute tumor sizes between caliper and MRI. On day 10, the mean tumor volumes measured with calipers and MRI were 254.78 mm³ and 462.93 mm³, respectively. On day 20, the mean tumor volumes measured with calipers and MRI were 574.49 mm³ and 1164.29 mm³, respectively.

Tumor volumes derived from caliper and MRI measurements were then normalized to the initial tumor volume from the day before treatment was initiated. When the normalized

caliper- and MRI-based measurements were tested for correlation (as in Figure 3 for absolute volumes), no correlation was found (Figure 4).

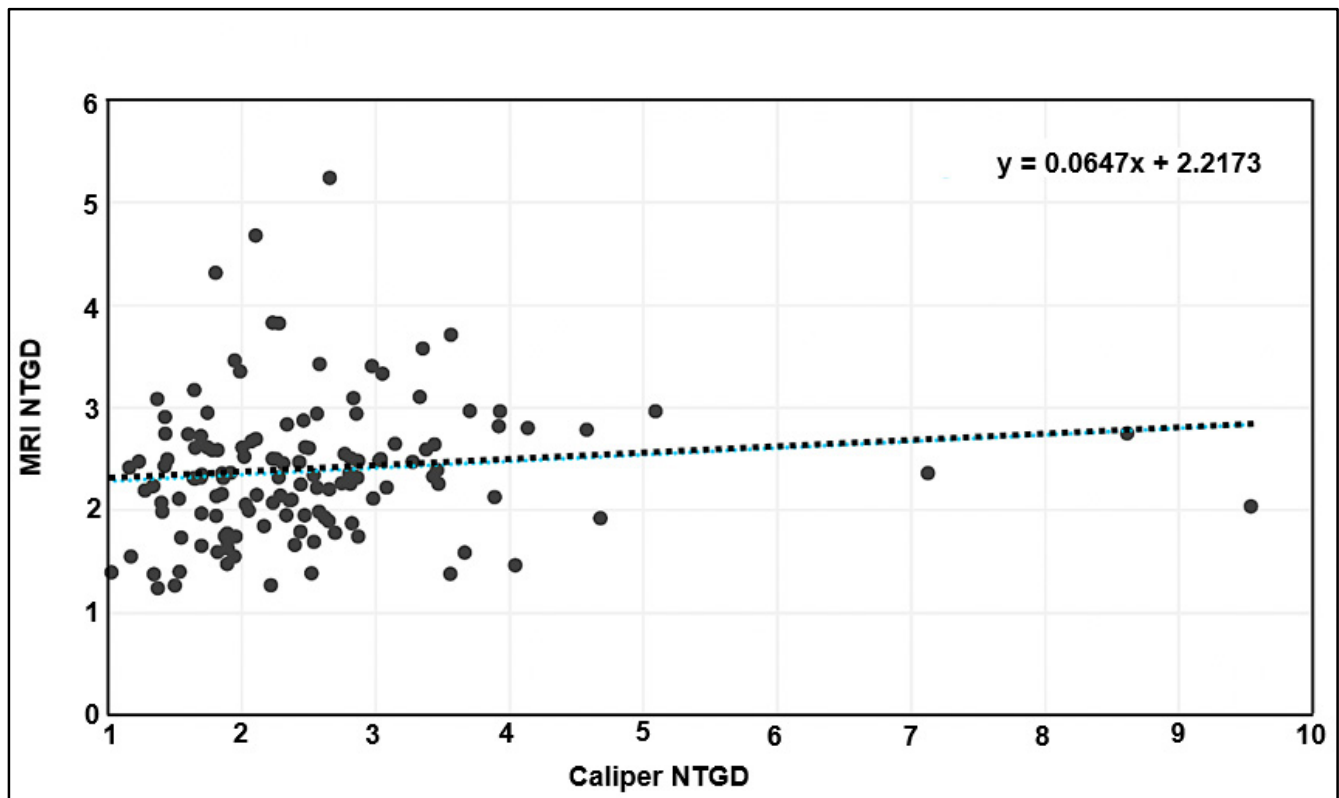


Figure 4. MRI versus caliper-based NTGD (pairwise). Each point represents a single tumor from the same mouse plotted pairwise—x-axis: that tumor’s caliper NTGD; y-axis: the MRI NTGD for the same tumor. (Each mouse has two tumors—right axilla and right inguinal mammary fat pad.) The dashed line shows the linear fit ($y = 0.0647x + 2.2173$), with Pearson $R^2 = 0.014$.

4. Discussion

Preclinical experiments evaluating oncologic interventions rely on longitudinal tumor-volume measurement. Although MRI is widely considered the ground truth for tumor-volume assessment, calipers are commonly used for subcutaneous tumors because they are inexpensive, fast, and easy to apply. The unanswered question is whether caliper measurements faithfully reflect MRI-based tumor volume. In this study, we demonstrate that caliper-derived tumor-volume measurements do not reflect MRI-based measurements using both normalized and non-normalized growth-delay analyses.

As expected, caliper measurements show a positive correlation with MRI tumor-volume measurements. Interestingly, as shown in Figure 3B,C, there is a more dramatic difference in tumor size, with MRI capturing a much larger volume than calipers. This is concerning, given that longitudinal caliper measurements increasingly depart from the ground truth over time. Additionally, when tumor volume is normalized, the positive correlation is lost (Figure 4), and a large fraction of caliper measurements show significantly higher size differences compared with MRI.

The discordance between caliper and MRI tumor-volume measurements demonstrates inaccuracy in the use of calipers. This warrants a closer look at whether the low cost of calipers justifies the loss in accuracy. This discrepancy could be a major contributor to the challenges of translating preclinical studies to clinical trials. Several studies fail in clinical trials, even when therapeutic benefit is demonstrated in mouse models [16]. Understandably, multiple factors contribute to this disconnect; however, there is a need to

focus on addressable factors. Lack of reproducibility in monitoring tumor-growth delay in preclinical models has been identified as a significant factor [17]. Therefore, standardization of MRI-based tumor-volume measurement in preclinical research should be considered. Alternatively, caliper measurements should be supplemented with histopathologic assessment at the time of sacrifice. Although this provides only a single time point, it is effectively the gold standard for response evaluation and is warranted, given the variability and uncertainty of caliper-derived volumes and subsequent response estimates.

This study has a few limitations. Although technicians A and B's caliper measurements were not significantly different—suggesting consistent skill—we do not know whether their technique represents the broader population of personnel proficient in caliper tumor-volume measurement. Thus, the significant differences observed between caliper and MRI could reflect suboptimal proficiency with calipers. This could be addressed by creating a standardized training and assessment protocol across research institutions and then repeating the MRI–caliper comparison only with technicians who pass the proficiency assessment. Also, this study did not address the cost–benefit analysis of MRI versus calipers. Undoubtedly, MRI is more expensive, but it is more precise, more reliable, and, importantly, provides a permanent record that can be re-examined in the future, if needed.

5. Conclusions

This study shows the inaccuracy of caliper-based measurements compared with MRI. Addressing this issue has the potential to improve the translatability of preclinical research and increase the efficiency of moving discoveries from bench to bedside.

Author Contributions: Conceptualization, I.B.M. and B.S.; methodology, I.B.M. and B.S.; software, I.O.; validation, I.O., and I.B.M.; formal analysis, I.B.M.; investigation, I.B.M., I.O.; resources, I.B.M.; data curation, I.O.; writing—original draft preparation, I.O.; writing—review and editing, I.B.M.; visualization, I.O.; supervision, I.B.M. and B.S.; project administration, I.B.M.; funding acquisition, I.B.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Animal Care and Use Committee (protocol: 17-214-ad02 EDR, 27 January 2022) of the University of Miami.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data can be obtained from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ImT	Immunotherapy
RT	Radiotherapy
NTGD	Normalized tumor-growth delay
BLI	Bioluminescence imaging
CT	Computed tomography
US	Ultrasound
MRI	Magnetic resonance imaging

References

1. Ireson, C.R.; Alavijeh, M.S.; Palmer, A.M.; Fowler, E.R.; Jones, H.J. The role of mouse tumour models in the discovery and development of anticancer drugs. *Br. J. Cancer* **2019**, *121*, 101–108. [[CrossRef](#)]
2. Tomayko, M.M.; Reynolds, C.P. Determination of subcutaneous tumor size in athymic (nude) mice. *Cancer Chemother. Pharmacol.* **1989**, *24*, 148–154. [[CrossRef](#)] [[PubMed](#)]
3. Jensen, M.M.; Jørgensen, J.T.; Binderup, T.; Kjaer, A. Tumor volume in subcutaneous mouse xenografts measured by microCT is more accurate and reproducible than determined by 18F-FDG-microPET or external caliper. *BMC Med. Imaging* **2008**, *8*, 16. [[CrossRef](#)] [[PubMed](#)]
4. Ayers, G.D.; McKinley, E.T.; Zhao, P.; Fritz, J.M.; Metry, R.E.; Deal, B.C.; Adlerz, K.M.; Coffey, R.J.; Manning, H.C. Volume of preclinical xenograft tumors is more accurately assessed by ultrasound imaging than manual caliper measurements. *J. Ultrasound Med.* **2010**, *29*, 891–901. [[CrossRef](#)] [[PubMed](#)]
5. Fiennes, A.G. Growth rate of human tumour xenografts measured in nude mice by in vivo cast modelling. *Br. J. Surg.* **1988**, *75*, 23–24. [[CrossRef](#)] [[PubMed](#)]
6. Hector, S.; Prehn, J.H. Apoptosis signaling proteins as prognostic biomarkers in colorectal cancer: A review. *Biochim. Biophys. Acta* **2009**, *1795*, 117–129. [[CrossRef](#)] [[PubMed](#)]
7. Ishimori, T.; Tatsumi, M.; Wahl, R.L. Tumor response assessment is more robust with sequential CT scanning than external caliper measurements. *Acad. Radiol.* **2005**, *12*, 776–781. [[CrossRef](#)] [[PubMed](#)]
8. Cheung, A.M.; Brown, A.S.; Hastie, L.A.; Cucevic, V.; Roy, M.; Lacefield, J.C.; Fenster, A.; Foster, F.S. Three-dimensional ultrasound biomicroscopy for xenograft growth analysis. *Ultrasound Med. Biol.* **2005**, *31*, 865–870. [[CrossRef](#)] [[PubMed](#)]
9. Graham, K.C.; Wirtzfeld, L.A.; MacKenzie, L.T.; Postenka, C.O.; Groom, A.C.; MacDonald, I.C.; Fenster, A.; Lacefield, J.C.; Chambers, A.F. Three-dimensional high-frequency ultrasound imaging for longitudinal evaluation of liver metastases in preclinical models. *Cancer Res.* **2005**, *65*, 5231–5237. [[CrossRef](#)] [[PubMed](#)]
10. Rojas, J.D.; Joiner, J.B.; Velasco, B.; Bautista, K.J.B.; Aji, A.M.; Moore, C.J.; Beaumont, N.J.; Pylyayeva-Gupta, Y.; Dayton, P.A.; Gessner, R.C.; et al. Validation of a combined ultrasound and bioluminescence imaging system with magnetic resonance imaging in orthotopic pancreatic murine tumors. *Sci. Rep.* **2022**, *12*, 102. [[CrossRef](#)] [[PubMed](#)]
11. He, Z.; Evelhoch, J.L.; Mohammad, R.M.; Adsay, N.V.; Pettit, G.R.; Vaitkevicius, V.K.; Sarkar, F.H. Magnetic resonance imaging to measure therapeutic response using an orthotopic model of human pancreatic cancer. *Pancreas* **2000**, *21*, 69–76. [[CrossRef](#)] [[PubMed](#)]
12. Mazurchuk, R.; Graves, D.; Raghavan, D. Magnetic resonance imaging of response to chemotherapy in orthotopic xenografts of human bladder cancer. *Clin. Cancer Res.* **1997**, *3*, 1635–1641. [[PubMed](#)]
13. Ramasawmy, R.; Johnson, S.P.; Roberts, T.A.; Stuckey, D.J.; David, A.L.; Pedley, R.B.; Lythgoe, M.F.; Siow, B.; Walker-Samuel, S. Monitoring the Growth of an Orthotopic Tumour Xenograft Model: Multi-Modal Imaging Assessment with Benchtop MRI (1T), High-Field MRI (9.4T), Ultrasound and Bioluminescence. *PLoS ONE* **2016**, *11*, e0156162. [[CrossRef](#)] [[PubMed](#)]
14. Takahashi, Y.; Yasui, T.; Minami, K.; Tamari, K.; Otani, K.; Seo, Y.; Isohashi, F.; Koizumi, M.; Ogawa, K. Radiation Enhances the Efficacy of Antitumor Immunotherapy with an Immunocomplex of Interleukin-2 and Its Monoclonal Antibody. *Anticancer Res.* **2017**, *37*, 6799–6806. [[CrossRef](#)] [[PubMed](#)]
15. Tesei, A.; Arienti, C.; Bossi, G.; Santi, S.; De Santis, I.; Bevilacqua, A.; Zanoni, M.; Pignatta, S.; Cortesi, M.; Zamagni, A.; et al. TP53 drives abscopal effect by secretion of senescence-associated molecular signals in non-small cell lung cancer. *J. Exp. Clin. Cancer Res.* **2021**, *40*, 89. [[CrossRef](#)] [[PubMed](#)]
16. Curran, T. Reproducibility of academic preclinical translational research: Lessons from the development of Hedgehog pathway inhibitors to treat cancer. *Open Biol.* **2018**, *8*, 180098. [[CrossRef](#)] [[PubMed](#)]
17. Sugar, E.; Pascoe, A.J.; Azad, N. Reporting of preclinical tumor-graft cancer therapeutic studies. *Cancer Biol. Ther.* **2012**, *13*, 1262–1268. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.