

# Trends in Subcutaneous Tumour Height and Impact on Measurement Accuracy

Daniel Brough, Hope Amos, Karl Turley and Jake Murkin

BioVolume Ltd, Witney Business & Innovation Centre, Witney, UK.

Cancer Informatics  
Volume 22: 1–14  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11769351231165181



**ABSTRACT:** Tumour volume is typically calculated using only length and width measurements, using width as a proxy for height in a 1:1 ratio. When tracking tumour growth over time, important morphological information and measurement accuracy is lost by ignoring height, which we show is a unique variable. Lengths, widths, and heights of 9522 subcutaneous tumours in mice were measured using 3D and thermal imaging. The average height:width ratio was found to be 1:3 proving that using width as a proxy for height overestimates tumour volume. Comparing volumes calculated with and without tumour height to the true volumes of excised tumours indeed showed that using the volume formula including height produced volumes 36X more accurate (based off of percentage difference). Monitoring the height:width relationship (prominence) across tumour growth curves indicated that prominence varied, and that height could change independent of width. Twelve cell lines were investigated individually; the scale of tumour prominence was cell line-dependent with relatively less prominent tumours (MC38, BL2, LL/2) and more prominent tumours (RENCA, HCT116) detected. Prominence trends across the growth cycle were also dependent on cell line; prominence was correlated with tumour growth in some cell lines (4T1, CT26, LNCaP), but not others (MC38, TC-1, LL/2). When pooled, invasive cell lines produced tumours that were significantly less prominent at volumes >1200 mm<sup>3</sup> compared to non-invasive cell lines ( $P < .001$ ). Modelling was used to show the impact of the increased accuracy gained by including height in volume calculations on several efficacy study outcomes. Variations in measurement accuracy contribute to experimental variation and irreproducibility of data, therefore we strongly advise researchers to measure height to improve accuracy in tumour studies.

**KEYWORDS:** Tumour biology, imaging, in vivo, metastasis, in silico, preclinical oncology, volume calculation, accuracy

**RECEIVED:** January 11, 2023. **ACCEPTED:** March 5, 2023.

**TYPE:** Original Research

**FUNDING:** The author(s) received no financial support for the research, authorship, and/or publication of this article.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: BioVolume Ltd (an operating company of Fuel 3D Technologies Limited) is producing the BioVolume units and software and claims financial competing interests on

the product. There are specific patents granted and filed for this technology or any part of it. BioVolume Ltd provided support in the form of salaries for authors but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. In vivo work and all measurements were carried out by BioVolume users who were not employed by BioVolume Ltd and who did not receive financial compensation.

**CORRESPONDING AUTHOR:** Daniel Brough, BioVolume Ltd, Witney Business & Innovation Centre, Windrush Industrial Park, Burford Road, Witney OX29 7DX, UK. Email: dan@fuel3d.com

## Introduction

Therapeutic efficacy of potential cancer therapies is commonly assessed by monitoring changes in subcutaneous tumour volume over time, by comparing tumour growth curves between control and treated groups of rodents. Accuracy is important both at small tumour volumes, when performing randomisation and determining the start of treatment and dosing, and at large volumes which are used when determining drug efficacy against controls, and to confirm if animal welfare points have been reached. Greater volume measurement accuracy can therefore give greater confidence in results with the potential to run shorter studies, or requiring fewer repeats to achieve the same level of statistical significance in results.

Important information about the tumour is lost when recording only tumour length and width over time, as is standard in preclinical studies. Subcutaneous tumours are inoculated below the dermis and grow outwards, protruding from the rodent body. On average, the tumour shape can be generalised as an ellipsoid where height and width are equal,<sup>1</sup> however the shape and prominence (height:width relationship) of subcutaneous tumours can be affected by several variables including injection technique, cell line, tumour type, and growth stage. Large subcutaneous tumours for example, are known to have a non-spherical, flattened shape.<sup>2</sup> Complex tumours may have multiple lobes,<sup>3</sup> and tumours where a necrotic core develops

may also collapse in on themselves, producing a 'donut' or 'cratered' shape.<sup>4</sup> Tumours with an aggressive, metastatic phenotype gain the ability to invade into other tissues during the transition to metastasis.<sup>5</sup> We might reasonably expect that this local invasiveness at the primary site would be detectable by tracking tumour prominence over time.

Tumour type and cell line also affect subcutaneous tumour shape. Lymphoma tumour cell lines form tumours with less stroma than other tumour types including carcinomas and produce characteristic 'diffuse' or 'flattened' subcutaneous tumours that can be distinguished from carcinomas by eye.<sup>6</sup> Anecdotally, technicians have reported that such diffuse, flat tumours are difficult to measure consistently using standard calliper techniques due to their poorly defined boundaries that are difficult to define by eye and surround with calliper blades.

The low-cost, high throughput method of choice for capturing tumour volume is manual measurement with callipers. The longest tumour length and width measurements (as judged by the user) are taken by placing calliper blades on the outside of the tumour bulk protruding from the mouse body. Calliper measurement is known to introduce bias as the direction to measure the longest length is subjective, especially in tumours with complex morphologies.<sup>5</sup> Calliper contact with the tumour can also cause squeezing of the tumour which leads to underestimation of the true volume.



Height is difficult to capture using callipers as the tumour base is obstructed by the mouse body, leading to larger coefficient of variation and instrumental error from height measurement.<sup>2</sup> Measurement of height also increases animal handling time when using callipers. Thus, methods to estimate tumour volume from length and width alone are used most commonly, including the LWW (Volume formula using length and width) formula which uses tumour width as a proxy for height in a 1:1 ratio to estimate tumour volume.<sup>7,8</sup> However, this formula causes overestimation of tumour volume as it does not take into account any changes in height which is smaller than width.<sup>5</sup> Different approaches have been suggested in order to achieve more accurate tumour volumes without measuring height<sup>2,8</sup> but these methods have not been widely adopted or implemented.

#### LWW Formula:

$$n \times \text{length} \times \text{width} \times \text{width}$$

where  $n$  is a constant, most commonly  $\pi/6$  (used here) or 0.5.

Alternative measurement devices offer increased accuracy and reduced bias including computerised tomography (CT)-,<sup>9</sup> fluorescence-,<sup>10</sup> Magnetic Resonance-,<sup>11</sup> and ultrasound imaging.<sup>12</sup> These techniques allow visualisation and measurement of the whole tumour in situ, however the main disadvantages; cost, low throughput, and significant welfare risks from anaesthesia and invasive techniques ensure that calliper measurement remains the standard for subcutaneous tumour measurement.

3D and thermal imaging (3D-TI) is a technique that captures and records tumour morphology for volume calculation.<sup>13</sup> In comparison to callipers, 3D-TI reduces user bias by using machine learning and computational algorithms to calculate length, width, and height measurements automatically. The variability in calliper tumour measurement methods is predicted to affect the rate of false negative results in efficacy studies,<sup>14</sup> and the tumour growth inhibition (TGI) drug efficacy metric<sup>15</sup> in comparison to 3D-TI. Height measurement allows 3D-TI to calculate tumour volume using the LWH (Volume formula using length, width and height) formula which has been shown to give more consistently accurate volumes in comparison to the LWW formula and 18 other tumour volume formulae.<sup>7</sup> The LWH formula has also been proven to improve accuracy in clinical studies.<sup>16</sup>

#### LWH Formula<sup>7</sup>:

$$\pi / 6 \times \text{length} \times \text{width} \times \text{height}$$

#### Aims

Including tumour height in volume calculations has been shown to produce more accurate tumour volumes than formulae using length and width dimensions alone. We aimed to use our anonymised global database of tumour scans taken by

3D-TI users to first define the average relationship between tumour height and width, then aimed to validate our findings by comparing LWH and LWW tumour volumes to the true volumes of excised tumours.

The second part of the investigation aimed to analyse the real-world impact of variations in tumour prominence in more detail, by breaking cell lines down into metastatic versus non-metastatic groups, and grouping by tumour types. To quantify the importance of measuring tumour height during studies, *in silico* modelling was used to simulate the effect on efficacy study outcome metrics (TGI, false positive rate, and false negative rate) when comparing LWW and LWH measurement formulae.

## Methods

### *In vivo data collection*

The 3D-TI device and system (BioVolume) and support were provided to 27 client organisations by BioVolume Ltd (previously Fuel 3D Technologies Limited). Subcutaneous tumour models were established in the mouse flank area from inoculated cell lines according to client organisation protocols. Training was provided by BioVolume Ltd over a series of sessions either in person or online. Mice were scruffed according to client animal handling protocols, and the tumour region was presented to the BioVolume silicon aperture for image capture (which takes ~0.5 second). Tumour length, width, and height were measured automatically by the 3D-TI device as previously described.<sup>13</sup>

Methods and welfare endpoints adhered to US or EU welfare and ethics rules depending on lab location. All animal care, lab work, calliper measurements, and 3D-TI scans were carried out by scientists in client organisations. Data were shared with BioVolume to use in an aggregated and anonymised way. Client companies and scientists did not have financial interests in BioVolume.

Maximum tumour volume limits are typically 1.2 to 1.5 cm diameter in the UK<sup>17</sup> and 2 cm diameter or ~2000 mm<sup>3</sup> in the US, therefore we looked at measurement accuracy up to ~2000 mm<sup>3</sup>.

### *Data processing*

*Global dataset.* Our global dataset of tumour scans included 39 758 scans from 119 cell lines and 156 users in 27 organisations.

*Evaluation dataset.* Scans in the global dataset were filtered to include data from evaluation studies only, as these studies were performed after user training was completed, with a high level of scan quality being maintained by users. This comprised 9522 scans from 16 cell lines and 46 users (Evaluation Dataset). Only established cell lines with no additional genetic modification were included. To better account for any non-linearity, a

**Table 1.** Metastatic profiles of subcutaneously inoculated tumour cell lines used when comparing cell lines.

CELL LINE	METASTATIC PROFILE	NO. 3D-TI SCANS	TUMOUR TYPE
MC38	Low <sup>21</sup>	1020	Murine colon adenocarcinoma
4T1	High <sup>22,23</sup>	481	Mammary carcinoma
TC-1	High <sup>24</sup>	419	Mouse lung epithelial carcinoma
LNCaP	None <sup>25</sup>	985	Human prostatic carcinoma
CT26	Low/none <sup>26,27</sup>	657	Colorectal carcinoma
RENCA	None <sup>28</sup>	607	Renal cell carcinoma

generalised additive model (GAM)<sup>18</sup> was applied to this dataset to better understand the relationship between height and width across different tumour volumes.

*Control dataset.* Scans in the Global Dataset were assigned to a Study and Group by users during scanning. We identified control groups in the Global Dataset using group names. Scans in other groups were filtered out, to reduce potential effects of treatments on tumour morphology and to enable like-for-like comparison between untreated tumour cell lines. Any ambiguously labelled groups were excluded and only immortalised cell lines were included. Forty-two cell lines and 9067 scans were present in this dataset (Control Dataset).

*Excised weights dataset.* The global dataset was filtered to remove any scans where excised tumour weight was not available. This dataset consisted of 299 3D-TI scans and 274 corresponding calliper measurements across 13 studies, 12 cell lines and 6 mouse strains.

*Modelling.* Testing confirmed that the residuals of all models were normally distributed, thereby meeting the assumptions of linear modelling. To verify that models described the data well, predicted tumour measurements were plotted against the actual tumour measurements for both models, a  $y=x$  line was plotted and used to assess model fit (all available in Supplemental Materials).

#### *Tumour prominence by cell line*

To begin the investigation into tumour height, an initial exploratory plot using the Evaluation Dataset scans was created to observe the height:width (prominence) relationship across all 16 cell lines. Tumour height was plotted against tumour width as measured by 3D-TI. A generalised additive model with integrated smoothness integration using the 'REML' method<sup>19,20</sup> was then fitted to all the Evaluation Data using the ggplot2 geom\_smooth function in R<sup>19</sup> with an intercept set through the origin to see if there was a linear relationship between the height and width, along with a 95% confidence interval associated with the prediction of the linear model.

Six immortalised cell lines in the Evaluation Dataset were chosen to study in greater detail: 4T1, CT26, LNCaP, MC38, RENCA and TC-1 (Table 1). They were amongst the most frequently measured tumour cell lines, and all represent carcinoma tumours. A linear model was fit (using stat\_smooth<sup>19</sup>) to each cell line to assess height:width relationship, along with a 95% confidence interval associated with the prediction of the linear model.

Following the initial exploratory plot and the detection of a non-linear height:width relationship, volume bins were used to group tumours by size for a new linear model. The lm function from the stats package in R was used.<sup>29</sup> This way prominence was directly compared across cell lines without the influence of tumour size, providing more confidence in the results that any differences would be down to the cell line. No intercept term was used and the global intercept was set to 0 since a tumour of 0 width should also have 0 height, which in turn meant that the fit would have to pass through the origin. Data were removed for cell lines and bins combinations that contained less than 3 datapoints as this was not enough data to create a linear model.

The model formula is outlined below:

$$\text{Height} = \text{Width} : \text{Tumour.Name} : \text{Bin}$$

Control group data was used in order to remove any effects that treatment may have had on prominence. The new linear model was then applied to the same immortalised cell lines (4T1, CT26, LNCaP, MC38, RENCA and TC-1) as previously mentioned, with the number of scans for each cell line outlined in Table 1. 95% confidence intervals associated with the slope estimates were also provided. A 2-way Analysis of variance (ANOVA)<sup>30,31</sup> was computed for unbalanced designs (type III) to test whether the variability between cell lines was greater than the variability within cell lines.

#### *Excised weights*

To calculate tumour weights, tumours were excised on the final measurement day. The true volume of excised tumours was derived by weighing the tumour, then volume was calculated in a 1:1 ratio (mg to mm<sup>3</sup>). The difference between calculated

**Table 2.** Cell lines used to compare prominence trends between metastatic and non-metastatic cell lines.

TUMOUR NAME	METASTASIS FROM SUBCUTANEOUS SITE	SCANS
HCT116	No <sup>32</sup>	188
LNCaP	No <sup>25</sup>	985
MCF7	No <sup>33</sup>	228
MiaPaCa2	No <sup>34</sup>	376
RENCA	No <sup>28</sup>	607
4T1	Yes <sup>22,23</sup>	481
B16F10	Yes <sup>24</sup>	465
H358	Yes <sup>35</sup>	243
LL/2	Yes <sup>36</sup>	176
TC-1	Yes <sup>24</sup>	419

volume and excised weight was calculated in a relative fashion (percentage difference) to compare the differences between the volumes obtained.

Using the excised weights dataset, prominence (height divided by width) was calculated from the 3D-TI measurement and matched with the corresponding calliper measurement where data were available, then plotted against the percentage difference from excised weight. A linear model was then fitted to the graph along with the error associated with the model.

#### *Prominence by metastatic profile*

Cell lines studied previously were classified by metastatic profile according to reports in the literature (Table 1). Only metastasis from tumours inoculated subcutaneously in mice was considered to be relevant to our study; metastasis from intravenous or orthotopic inoculation was discounted. Cell lines where a definitive classifier could not be found in the literature were excluded. CT26 was both reported as non-metastatic,<sup>26</sup> and to have some incidences of metastasis from subcutaneous tumours <2000 mm<sup>3</sup> which increased over time.<sup>27</sup> MC38 was characterised as having a low rate of metastasis from subcutaneous sites because metastasis rate was <50% although final tumour volumes were much larger than today's permitted maximum volume endpoints.<sup>21</sup>

3D-TI data from the Control Dataset were sorted into 'Metastatic' and 'Non-Metastatic' groups in the same way. Data were filtered such that any cell lines with uncertainty surrounding their metastatic nature were excluded for example, conflicting results in the literature. Cell lines with <150 tumour scans were also excluded, leaving 5 metastatic cell lines (1771 scans) and 5 non-metastatic cell lines (2380 scans) (Table 2). Data from this filtered dataset were plotted and 2 linear models were fitted, with 95% confidence intervals also being provided.

The model formulae are outlined below:

$$Height = Width : Metastatic : Bin$$

$$Height = Width : Tumour.Name : Bin$$

This was followed by computation of a 2-way ANOVA<sup>31</sup> for unbalanced designs (type III) and a Tukey's honest significant differences test. A general linear hypothesis test in the Multcomp R package<sup>37</sup> was then also conducted to test for significant differences in slope estimates of the model between the 2 groups for each of the individual volume bins.

#### *Prominence by tumour type*

To investigate whether cell lines of the same tumour type displayed similar prominence characteristics, cell lines within different groups from the Control Dataset were compared - these subgroups were composed of carcinomas,<sup>22</sup> gliomas,<sup>1</sup> melanomas,<sup>4</sup> leukaemias,<sup>3</sup> and lymphomas.<sup>2</sup> In order to give a fair representation of each cell type, cell lines with less than 9 datapoints in total were excluded. The control group data for carcinomas versus non-carcinomas were fed into a height:width linear model accounting for carcinoma status (yes or no) and each volume bin.

The model formula is outlined below:

$$Height = Width : Carcinoma : Bin$$

A 2-way ANOVA<sup>31</sup> for unbalanced designs (type III) and a Tukey's honest significant differences test were computed. A general linear hypothesis test in the Multcomp R package<sup>37</sup> was then used to test for significant differences between each of the volume bins.

#### *Generating the in silico efficacy dataset*

To outline the impact that failure to measure height could have on an efficacy result, the width-height models using volume bins described in the 'Tumour Prominence by Cell Line' section were used to create an artificial dataset for 4T1, CT26, LNCaP, MC38, RENCA and TC-1. Synthetic data were generated with the aim of removing any unwanted sources of variability that come with the aggregated data such as measurements coming from different studies and laboratories and to also generate a larger sample size. To consider cell line-specific growth rates, real data were fed into separate linear models for the length and width tumour dimensions, accounting for change in growth on a daily basis and creating an intercept for each study since not all studies started at the same volume (only studies with greater than 3 measurement sessions were used).

The model formulae are outlined below:

$$Length = day : Tumour.Name + Study.Name$$

$$\text{Width} = \text{day}:\text{Tumour.Name} + \text{Study.Name}$$

The obtained growth rates were then multiplied by a list containing the unique days within the synthetic study (multiples of 3 starting from 3 through to 90). Volumes were calculated using the LWW formula and a maximum volume limit of 2500 mm<sup>3</sup> was set. From these data, 10 synthetic rodents were generated and initial tumour volumes were offset by applying the same standard deviation to length and width as observed within the control group data for days and studies where the LWW volume was less than 120 mm<sup>3</sup>.

The individual rodent growth rates for length and width (gradients) were varied by a standard deviation equal to the slope error observed when calculating the growth rate coefficients for each cell line. The LWW volume for each rodent was then recalculated and results were generated. Lengths and widths that were less than 0 were excluded from the dataset (which had occurred due to the initial volume offset). The mean for each cell line on each day was calculated at this point. The data were then duplicated to create a treated group with identical starting conditions, with ‘treatment’ applied on the first day after the group mean exceeded 80 mm<sup>3</sup>. ‘Treatment’ was simulated by multiplying individual length and width dimensions of the treated group tumours by a defined treatment strength (from 0.5 to 0.9), with 0.5 being the strongest treatment level and 0.9 being the weakest. This had the effect of reducing the length and width growth rates by  $100 \times (1 - \text{treatment strength})\%$ , for example; at a treatment level of 0.9 the growth rate would be reduced by  $100 \times (1 - 0.9) = 10\%$  for the treated group beginning on the measurement session after first treatment was applied. Within this step, tumour heights were predicted by fitting data to the linear model for the cell lines and the LWH volume was computed. This step was repeated across 41 varying levels of treatment strength. This process was then repeated 250 times after generating 250 sets of 20 animals (10 control, 10 treated).

#### *Using the in silico dataset to calculate tumour growth inhibition*

The ratio of final volume to treated day volume was calculated for each rodent for LWW and LWH volumes. For each iteration, cell line, volume type, and efficacy level the tumour growth inhibition (TGI) was computed by dividing the mean rodent ratio for the treated group by that of the control group (group sizes were both equal). To show this graphically, the mean TGI was computed across the iterations and plotted with 95% confidence intervals against the efficacy level for each volume type and cell line. To calculate the number of successful results, a conservative TGI threshold of 0.45 was used and applied such that any result under 0.45 would be considered to have shown effectiveness of treatment and any result above 0.45 would show ineffective treatment. The LWH volume was taken to be the ground truth in this scenario as it was established to be the

most accurate volume formula because of the additional height measurement. The number of successful trials were calculated using the TGI cut-off (<0.45) and the results for each volume type were compared to each other. LWW error rates were calculated by comparing the number of successful trials in comparison to LWH for each cell line. For example, if LWW had a greater number of successful trials than the LWH standard then this would indicate false positive results in some LWW trials.

## Results

### *The tumour height:width relationship*

Plotting tumour height against width clearly showed that width is not a good proxy for tumour height, with none of the 9522 tumour height values being equal to or exceeding their corresponding width measurements (Figure 1A). Height was roughly one-third of the width on average: Height (mean) =  $(0.325 \pm 0.001) \times \text{Width (mean)}$ . The heteroskedastic nature of the plot means that as width increases, so do the range of possible height values for a given width. This reveals the unique and independent behaviour of tumour height, further outlining why using width as a proxy for height is not an accurate approach when calculating tumour volume. The next step was to focus on the height:width relationship for individual cell lines.

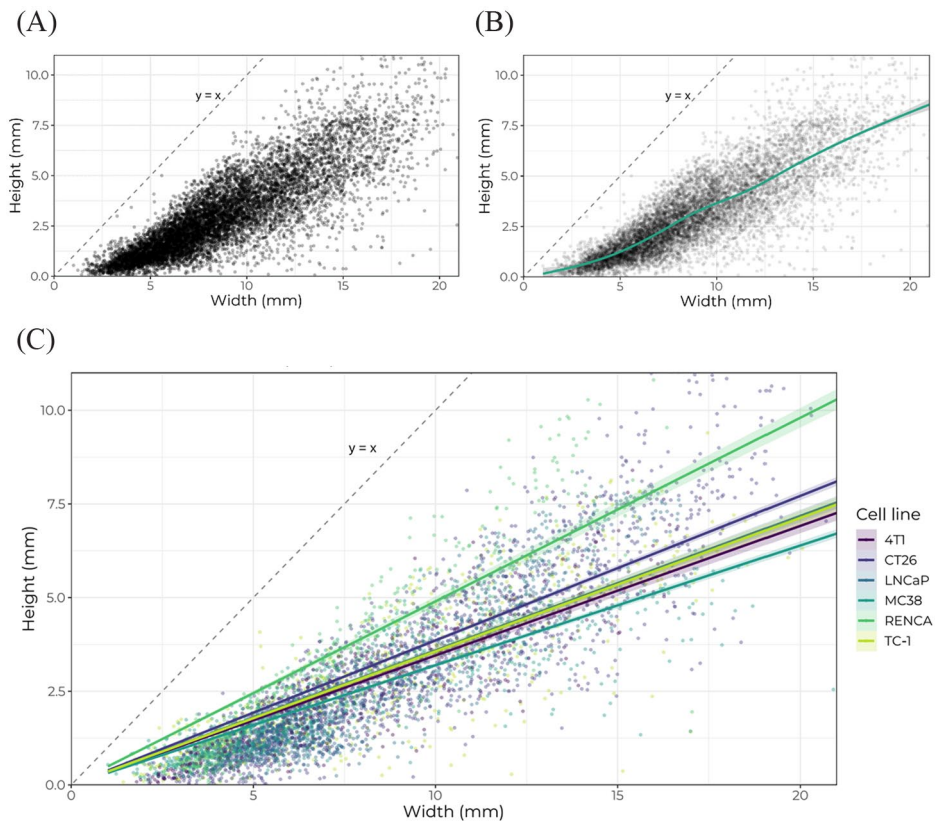
### *Tumour prominence by cell line*

Data for 6 cell lines (4T1, CT26, LNCaP, MC38, RENCA and TC-1) were plotted to investigate any unique cell line characteristics (Figure 1C). The tumours from different cell lines varied in prominence with respect to each other; RENCA had the greatest prominence (height/width) of the 6 tumours, with MC38 forming the least prominent (flattest) tumours. The linear model appeared to underfit by a large margin at smaller widths, suggesting a non-linear distribution of the data.

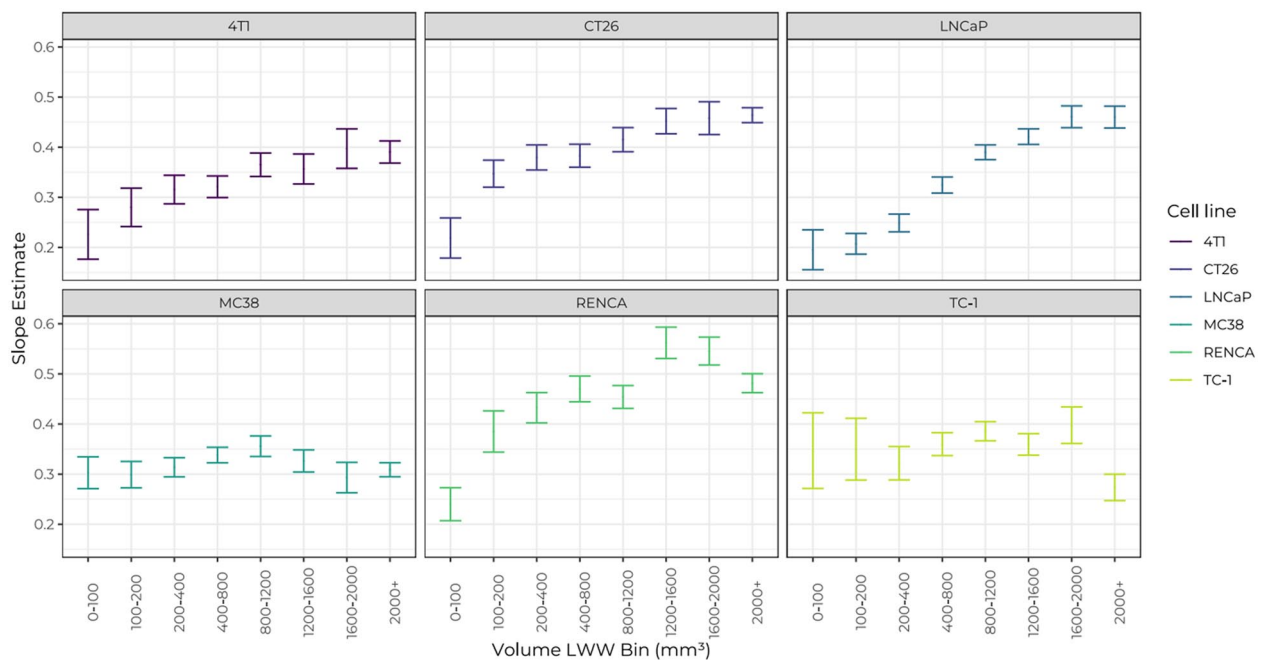
To confirm this hypothesis, a generalised additive model was fitted to the evaluation dataset (shown in Figure 1A) which, following observation of the fitted GAM slope, suggested a non-linear height:width relationship at smaller volumes (Figure 1B). This indicates that on average tumours were flattest (least prominent) at the earliest stages of growth, below widths of 8 mm.

### *Prominence trends across increasing tumour volumes*

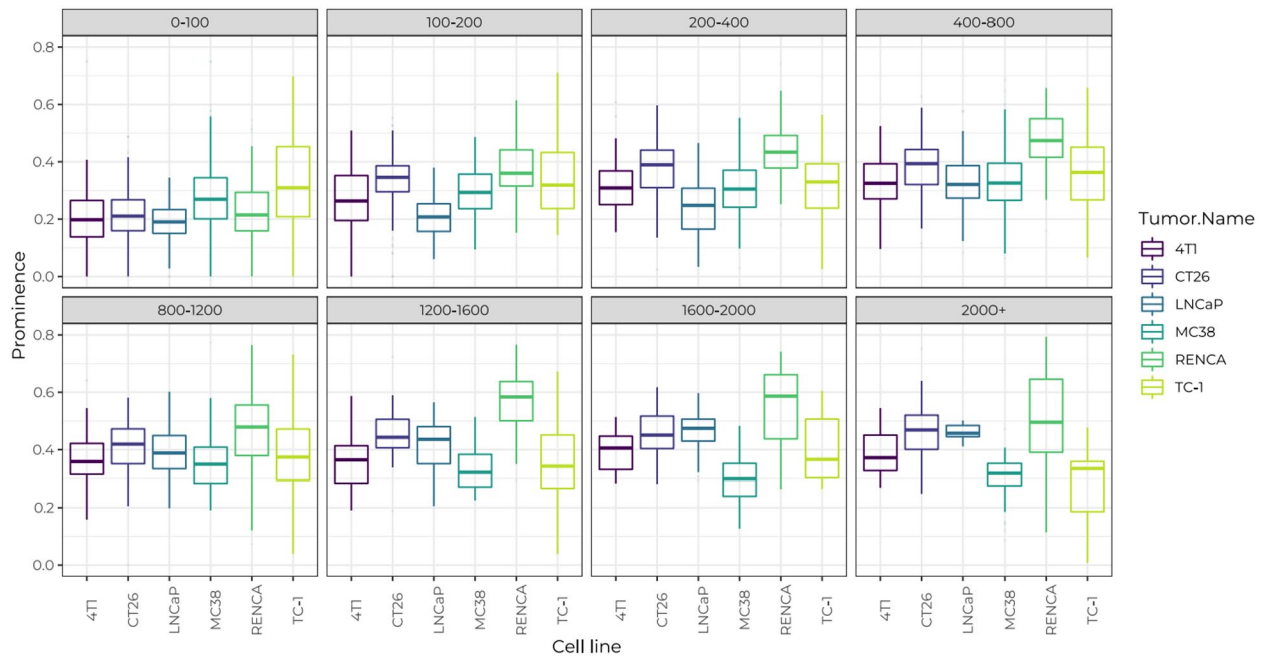
A new model introducing volume bins was applied to individual cell line data to determine if this non-linear height:width relationship was cell line-specific (Figure 2). The different trends for each of the cell lines show that the height:width ratio (prominence) can change as a tumour grows and that cell lines exhibit their own unique distributions. 4T1, CT26, LNCaP and RENCA tumour volume increase correlated with an



**Figure 1.** Investigation of subcutaneous tumour prominence. (A) Height versus width,  $n=9522$  3D-TI scans (Evaluation Dataset). Axes capped at (11 and 21 mm). (B) Generalised additive model with integrated smoothness integration using the 'REML' method<sup>19,20</sup> (green line) plotted on Figure 1A data, along with a 95% confidence interval associated with the prediction of the linear model (shaded). (C) Height versus width for subcutaneous tumours of 6 cell lines (4T1, CT26, LNCaP, MC38, RENCA, TC-1.  $N=5342$  total). Trend lines were fitted using a linear model for each cell line. Shaded line region shows the 95% confidence level associated with the prediction of the linear model.  $y=x$  shown as dashed line on each plot for reference.



**Figure 2.** Tumour prominence by cell line. Slope estimates of the Height=Width:Tumour.Name:Bin model for cell lines 4T1, CT26, LNCaP, MC38, RENCA and TC-1. Error bars represent the 95% confidence intervals of the slope estimates.



**Figure 3.** Prominence value distributions for each cell line. Distributions are shown as box plots separated out by LWV volume bin ( $\text{mm}^3$ ).

increase in prominence up to  $\sim 1200 \text{ mm}^3$ , however MC38 and TC-1 started off as relatively more prominent tumours but the height:width relationship then stayed relatively stable over the same range. RENCA tumours exhibited the fastest increase in height relative to width up to a peak at 1200 to  $1600 \text{ mm}^3$ . This varying ratio of height:width within a given cell line could impact efficacy results, as key height differences between groups could go unnoticed when only measuring length and width.

The individual distributions of tumour prominence by cell line and bin size were plotted to compare differences between cell lines (Figure 3). Two-way ANOVA yielded a  $P$ -value of  $P < .0001$  across the cell lines, volume bins and their interactions, showing that the variability across cell lines was significantly greater than the variability within them. Cell line pairs were also compared by ANOVA (Table 3) with statistical difference detected between cell lines such as CT26-4T1, but not for other pairings such as LNCaP-4T1. This indicated that tumour prominence is dependent on tumour cell line.

4T1, LNCaP and MC38 tumours were not significantly different to each other. TC-1 and CT26 were not significantly different to each other. RENCA tumours were significantly different to almost all other cell lines (Table 3) which was unsurprising as they had a higher median prominence than all other cell lines for volumes greater than  $100 \text{ mm}^3$ , indicating that RENCA tumours had the most prominent morphology (Figure 2).

#### Investigating metastatic cell lines

We hypothesised that metastatic tumours could affect accuracy of measurements due to aggressive growth inwards at the

subcutaneous inoculation site which might not be detected by callipers. Cell lines in the Control Dataset were categorised by metastatic profile. The pooled metastatic cell lines had smaller slope estimate values (less prominent morphology) at larger volumes in comparison to non-metastatic cell lines (Figure 4A). A Tukey's Honest significant differences test on a fitted ANOVA concluded that there was no significant difference between the pooled metastatic and non-metastatic groups ( $P$  value = .5142). However, when metastatic and non-metastatic cell lines were split into volume bins and compared, results showed that there were significant differences in metastatic and non-metastatic tumours 100 to  $200 \text{ mm}^3$  in size, and at volumes  $> 1200 \text{ mm}^3$  (Table 4). Splitting results by cell line revealed different prominence trends within the metastatic and non-metastatic groups, indicating that metastatic potential was not the only variable affecting tumour prominence (Figure 4B).

#### Investigation of tumour type

Analysis of trends within and across tumour type groups was hindered by a lack of different cell lines of each tumour type, so carcinoma cell lines were compared against pooled non-carcinoma cell lines (Figure 5). The tumour prominence trends between these groups were significantly different overall ( $P < .0001$ , Table 5), but more data on non-carcinoma cell lines should be collected before drawing any conclusions due to the non-carcinoma dataset being heavily comprised of B16F10 and BL2 measurements (845 from 1396 non carcinoma measurements, more on this is available in the Supplemental Text, Table S2). Specifically, carcinomas differed in prominence from non-carcinoma tumours between 100 and  $1600 \text{ mm}^3$ . Tumours from the single lymphoma cell line (BL2) appeared

**Table 3.** Comparison of cell line tumour prominence.

	4T1	CT26	LNCaP	MC38	RENCA	TC-1
4T1	x	***	ns	ns	***	***
CT26	***	x	***	***	**	ns
LNCaP	ns	***	x	***	***	***
MC38	ns	***	***	x	***	***
RENCA	***	**	***	***	x	ns
TC-1	***	ns	***	***	ns	x

Tukey honest significant differences test computed using fitted ANOVA to give multiple pairwise-comparison between cell line group means. Significance indicated:  $P > .05$  ns,  $P \leq .005$ \*\*,  $P \leq .001$ \*\*\*.

flatter than average (data not shown), however more lymphoma cell lines need to be tested to confirm whether this trend is common to all lymphoma cell lines.

#### *Calliper versus 3D-TI accuracy*

Tumour prominence was compared to the accuracy of each measurement device in Figure 6. Excised tumour volumes were used to compare the accuracy of calliper measurement using the standard LWW formula, and 3D-TI using the LWH formula. The 3D-TI LWH volumes obtained a mean difference from excised tumour volume of  $-1.4\% \pm 4.2\%$  whereas the LWW calliper volumes had a mean percentage difference of  $50.5\% \pm 4.7\%$  away from the excised weights, a difference 36X larger. 3D-TI volumes remained accurate at all levels of tumour prominence whereas the LWW calliper volumes became less accurate as prominence decreased. This was likely due to the overestimation of height by the LWW formula as previously shown.

#### *Simulating the effect of height measurement on efficacy study results*

Simulated growth curves for LWW volumes were plotted so that the characteristics of 6 cell lines could be studied in more detail (Figure 7A). As expected, these simulated cell lines had different 'growth' rates, with RENCA growing at the quickest rate and LNCaP growing at the slowest rate. Rodent tumours were synthetically generated and the average synthetic rodent growth curves for LWW and LWH were plotted for each cell line (10 synthetic rodents per group, LNCaP example shown in Figure 7B). As expected, the LWH plots produced similar overall growth kinetics albeit at comparatively lower volumes than LWW.

These LWW errors came in the form of false positives or negatives dependent on the cell line and treatment level. A trial was considered a success when the TGI score was less than 0.45, and the relative percentage (and exact number) of false negatives and positives was determined for each cell line (Figure 8B). For LNCaP, more than 1 in 11 LWW simulated trials resulted in a false negative (9.3%) and for TC-1 more

than 1 in 10 trials resulted in a false positive (10.7%) when compared to LWH. These results show that failure to measure tumour height can negatively affect conclusions from a trial.

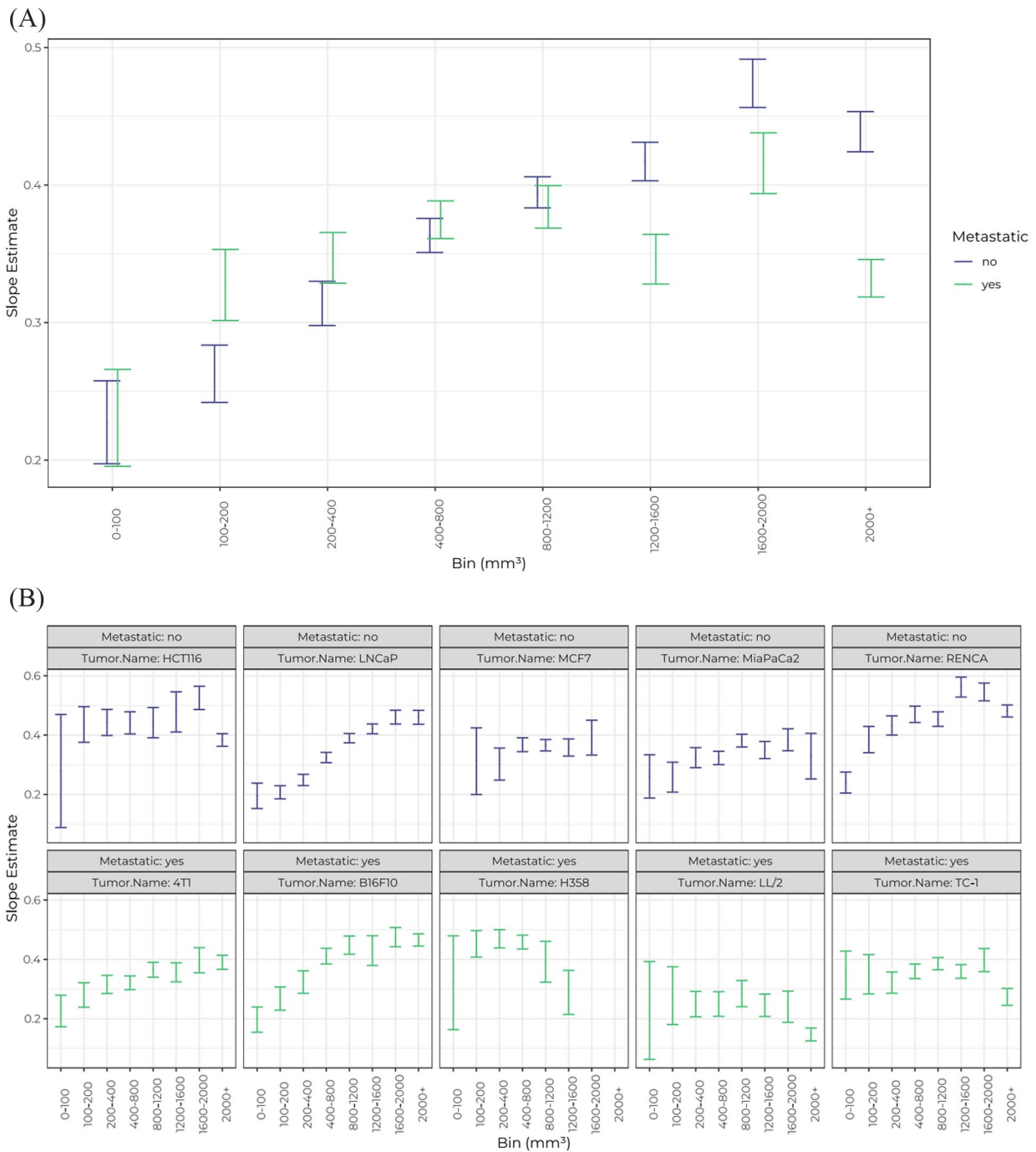
### Discussion

Despite the current standard to use tumour width as a proxy for height in volume calculations, our analysis of 9522 tumour scans showed that this is a poor approximation; tumour height is closer to one-third of the tumour width on average. We found that tumour prominence changes throughout the growth cycle and furthermore this relationship is cell line-dependent, therefore methods using width as a proxy for height, or which approximate tumour height mathematically using a constant are not sufficient to accurately estimate tumour volume.

Comparison of 3D-TI LWH and calliper LWW volume measurements to excised tumour weights further confirmed that measurement using 3 dimensions is needed for greatest volumetric accuracy. The discrepancy between measurement methods increased as tumour prominence decreased, and we found that tumour prominence is lowest (tumours are flattest) at small volumes. This finding is particularly relevant for randomisation which typically occurs at tumour volumes between 75 and 150 mm<sup>3</sup>, a range that can be affected by the presence of hydrogels such as Matrigel, scarring, or inflammation at the site of tumour inoculation. Accuracy at small volumes is therefore crucial in order to confirm significant growth up until the point of randomisation, ensuring the mass being measured is indeed an established tumour.

Determining the treatment window accurately at small volumes is also important during bioavailability and toxicity studies which require delivery of a therapeutic agent via the blood; the tumour must be large enough to be reliably established with a blood supply so that the drug can be delivered, but not too large that the optimum window for treatment has passed. Measuring height allows for greater accuracy when determining this point. 3D-TI not only provides more accurate results, but as Brough et al have previously shown, the automatic measurement feature reduces bias.<sup>15</sup> Greater confidence at such small volumes is needed for correct set up of experiments from the very beginning, and to ensure data are reproducible.





**Figure 4.** Slope estimates representing tumour prominence (Height divided by Width) for different tumour volumes. Error bars represent the 95% confidence intervals of the slope estimates. Metastatic (purple) and non-metastatic (green) groups and cell lines are indicated. (A) Slope estimates for metastatic (n=1769) and non-metastatic (n=2380) tumours. (B) Individual slope estimates for 10 tumour cell lines. Cell lines are labelled as metastatic or non-metastatic.

Six carcinoma cell lines were investigated individually, and the relationship between prominence and tumour size varied for each. The 6 cell lines had high metastatic, low metastatic, or non-metastatic profiles (2 of each) but results did not reveal significant differences in prominence correlated to metastatic profile. This is likely due to the large variation of tumour

morphology and growth kinetics found between different cell lines. Splitting the complete Control Dataset by tumour type revealed significant differences between carcinoma and non-carcinoma tumours however further analysis of a larger range of non-carcinoma cell lines is needed for more meaningful comparison. Two lymphoma cell lines did display below

average prominence, however more lymphoma cell lines need to be studied to confirm if this effect is true for all subcutaneous lymphoma models.

Investigating a predicted relationship between invasive cell lines and tumour prominence further confirmed significant differences in prominence trends when individual cell lines within metastatic and non-metastatic groups were compared. Although best efforts were made to categorise cell lines' metastatic potential based on available literature, reported experimental and data reporting methods varied, so it was difficult to definitively exclude cell lines as 'not metastatic

**Table 4.** Comparison of metastatic and non-metastatic tumour prominence across different volume ranges.

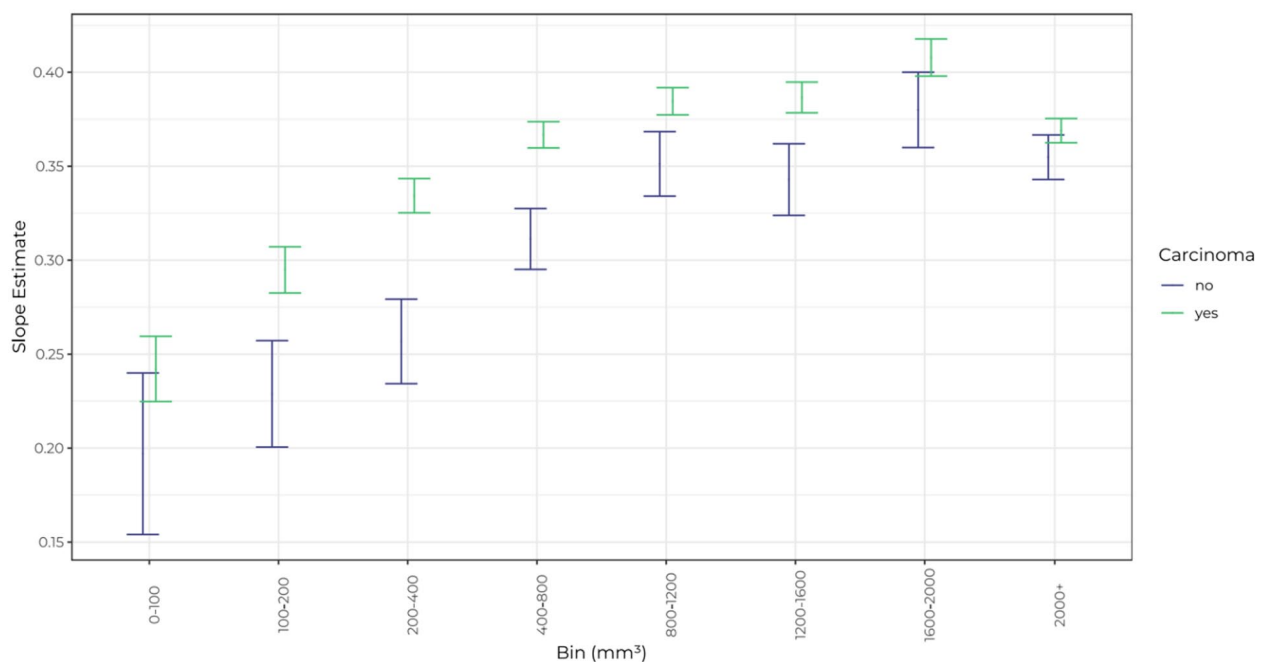
VOLUME (MM <sup>3</sup> )	P VALUE	SIGNIFICANCE
0-100	1.0000	ns
100-200	.0011	**
200-400	.0618	ns
400-800	.8700	ns
800-1200	.9263	ns
1200-1600	<.0001	***
1600-2000	.0004	***
2000+	<.0001	***

A general linear hypothesis test was computed to test for significance between each of the metastatic and non-metastatic volume bins. P values rounded to 4 dp. Significance indicated:  $P > .05$  ns,  $P \leq .005$ \*\*,  $P \leq .001$ \*\*\*.

from subcutaneous sites'.<sup>36</sup> We pooled data into metastatic and non-metastatic groups to provide a large sample size, and the results showed that overall, metastatic-type tumours were significantly less prominent (flatter) than non-metastatic tumours between volumes 100 and 200 mm<sup>3</sup>, and at volumes >1200 mm<sup>3</sup>. We predict that at large volumes, this observation could be correlated to the switch to an invasive phenotype as tumour cells begin to invade below the basement membrane and spread to other tissues. We suggest that tracking subcutaneous tumour prominence could therefore be investigated as an early warning marker for metastasis; this would require further study as trends would likely be cell line specific.

Most tumour studies last for weeks and therefore follow tumours across a range of volumes as they grow and/or regress. As tumour prominence is dependent on cell line and tumour size, volumetric accuracy is affected if the LWW formula is used, but the effect is not uniform. For cell lines that produce subcutaneous tumours of constant prominence throughout the growth cycle like MC38 and LL/2, the LWW formula will consistently overestimate volumes. For others like TC-1, RENCA, and H358 where prominence increases then decreases as tumours grow, accuracy will increase, then decrease in correlation with tumour growth.

When using cell lines with similar prominence profiles to LNCaP, B16F10 and CT26 that increase in prominence at larger volumes, accuracy when using LWW will be at its lowest at small volumes, increasing with tumour size. In particular, tumour models that follow this pattern are more likely to lead to more variable results in efficacy studies, as a decrease in measurement accuracy correlated with decreased size could



**Figure 5.** Slope estimates representing tumour prominence (Height divided by Width) for different carcinoma and non-carcinoma tumour volume ranges. Error bars represent the 95% confidence interval of the slope estimates. Number of scans: N = 7652 (carcinomas) and N = 1396 (non-carcinomas).

mean that measurement of tumour regression is less accurate than measurement of tumour growth. Lower accuracy at smaller volumes also impacts randomisation and dosing start points as previously discussed. These variations in accuracy contribute to experimental variation and irreproducibility of data. By using modelling to look at the effects of measurement inaccuracy caused by using the LWW formula versus LWH, we found that the failure to measure tumour height can result in more false positive or negative results. False positive results represent ineffective drugs being brought to clinical trial, wasting time and money in the process. Conversely, false negative

results represent missing a successful treatment effect and disregarding an effective drug candidate.

### Future Considerations

Callipers remain the most commonly used tool for tumour measurement but are rarely used to measure tumour height due to difficulty of the technique causing measurement variability even higher than when measuring length and width.<sup>2,38</sup> 3D-TI can be used to measure tumour height quickly and accurately, without reagents or anaesthesia so is an attractive alternative to current measurement devices to achieve accurate tumour volumes. Because all tumour measurements are captured from a single scan, measuring 3 dimensions is no slower than measuring 2, in contrast to measurements with callipers, where every additional measurement increases the measurement session and rodent handling time.

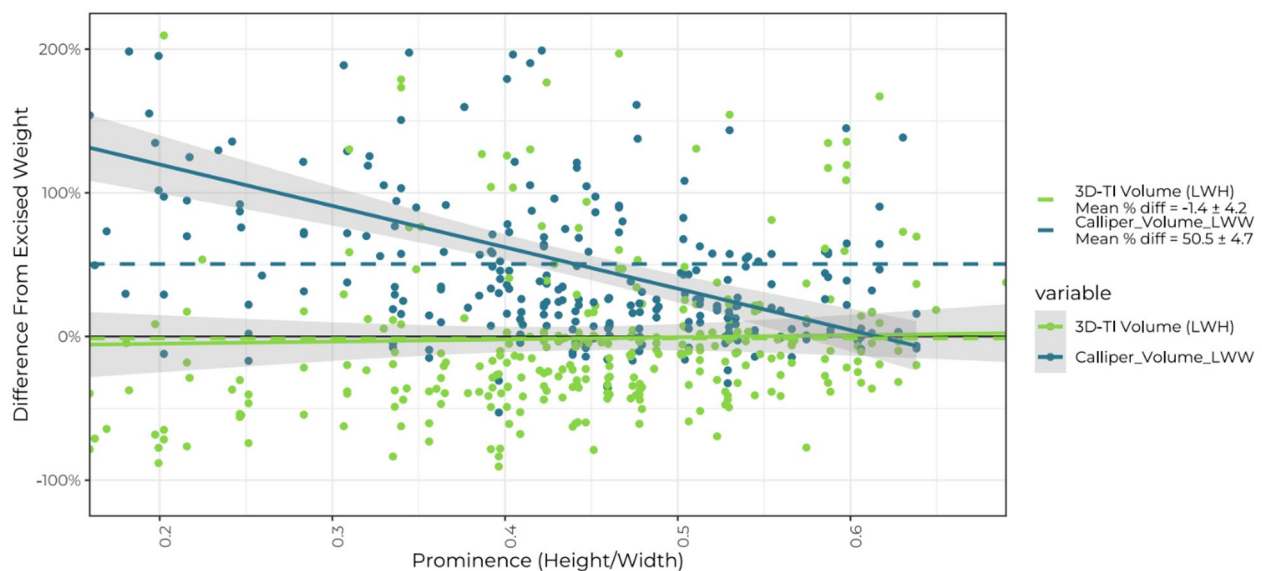
Because 3D-TI reconstructs whole tumours digitally as a 3D model, even greater accuracy could be achieved by calculating true volume and area. Although the 3D-TI, Magnetic resonance imaging (MRI), ultrasound, and other imaging technologies are ready for these innovations, the cancer research community is wary of moving too far away from the LWW formula upon which current welfare guidelines, dosing schemes, and randomisation points are based. Conversion models exist to solve these issues, however more support at institutional levels, time, and funding are needed to implement new technologies and methods.

We have shown that many tumour cell lines vary in prominence as subcutaneous tumour volume changes, and that height is a unique parameter independent of tumour width. For that

**Table 5.** Comparison of carcinoma and non-carcinoma tumour prominence across different volume ranges.

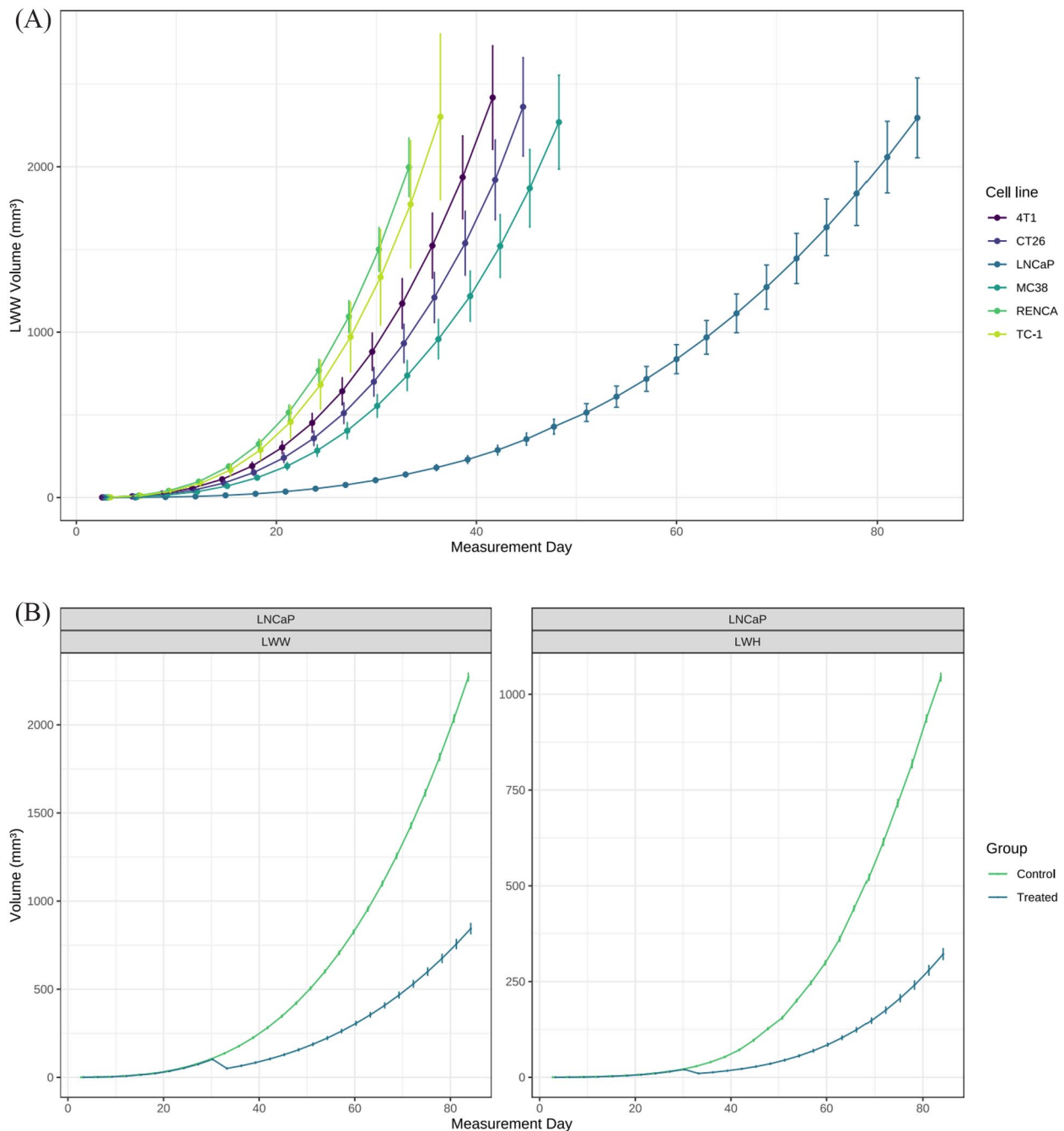
VOLUME BIN (MM <sup>3</sup> )	P VALUE	SIGNIFICANCE
0-100	.3710	ns
100-200	.0002	***
200-400	.0000	***
400-800	.0000	***
800-1200	.0036	**
1200-1600	.0003	***
1600-2000	.1100	ns
2000+	.2840	ns

A general linear hypothesis test was computed to test for significance between each of the carcinoma and non-carcinoma volume bins. P values rounded to 4 dp. Significance indicated:  $P > .05$  ns,  $P \leq .005$  \*\*,  $P \leq .001$  \*\*\*.



**Figure 6.** Accuracy of in vivo tumour measurement.

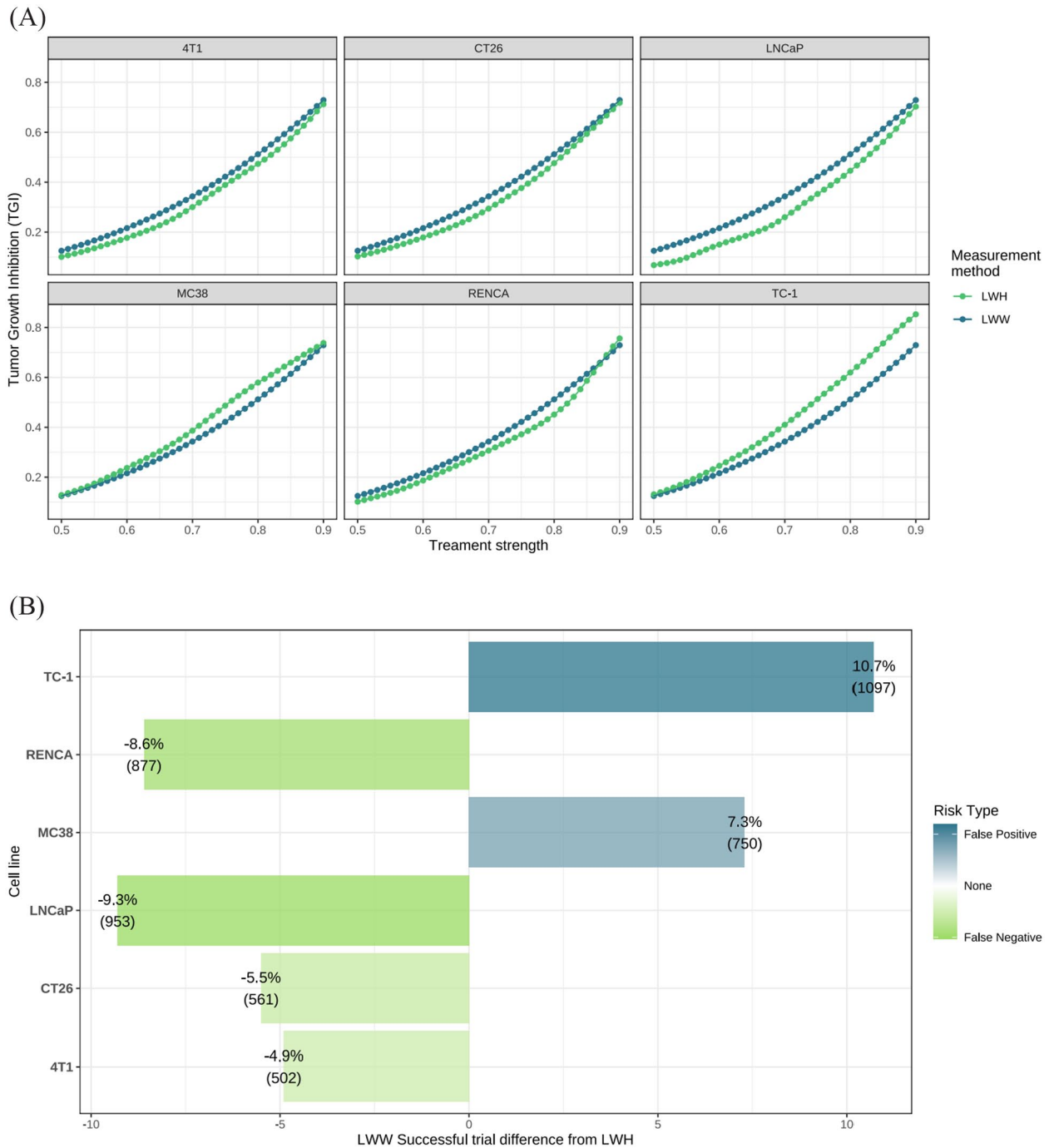
Relative percentage accuracy was calculated by comparing in vivo tumour measurements to excised tumour volumes. In vivo volumes were calculated using 3D-TI LWH (green) and calliper LWW (blue) methods. Data points are ordered by prominence (Height/Width) values as recorded by 3D-TI. Trend lines (solid) were fitted using a linear model for each variable; shaded regions represent 95% CIs. Mean percentage difference for each method is shown as a dotted line ( $-1.4\% \pm 4.2\%$  for 3D-TI LWH,  $50.5\% \pm 4.7\%$  for calliper LWW). Y axes cropped at  $-125\%$  and  $225\%$ .  $N=299$  for 3D-TI,  $N=274$  for callipers.



**Figure 7.** Simulated growth curves. (A) Simulated LWW growth rates for each cell line on each measurement day (using length and width growth rates modelled from real 3D-TI scans) with their associated error represented by 95% CIs. (B) Average simulated growth rates for LNCaP, for each group (Control or Treated), on each measurement day. LWW (left) and LWH (right). Associated error represented by 95% CIs. TGI was calculated from the rodent ratios in the control and treated groups for each measurement method to determine treatment effect. The average TGI score from 250 repeats across 41 different treatment strengths was plotted for each of the cell lines and measurement methods (Figure 8A). TGI score was dependent on volume formula and cell line, as well as treatment strength. Therefore, different types of errors may be observed depending on the TGI threshold for effective treatment and the cell line concerned. For example, the LWH TGI curves for LNCaP, CT26, 4T1 and RENCA (for the most part) are lower than the LWW curve, meaning that TGI scores obtained using LWH will be smaller and therefore LWH could more easily detect a treatment effect in these cell lines. On the other hand, MC38 and TC-1 (greater than a certain treatment strength of  $\approx 0.64$ ) show the opposite effect, meaning that the use of LWW could incorrectly show a treatment effect when there is none. We conclude that the effect of including height in volume calculation on TGI is cell line-dependent.

reason, we strongly advise against using the LWW formula, or other formulae that exclude tumour height. In our opinion, the best practice for accurate tumour volume measurements is by

including height, using the LWH formula. Tumour height is an independent variable that is largely ignored by scientists performing *in vivo* trials using callipers, and due to the



**Figure 8.** Simulated treatment effectiveness and error rates. (A) Average (mean) TGI values from 250 repeats per treatment level are plotted for LWW (blue) and LWH (green) measurement methods. Twenty simulated animals were observed on each trial; 10 control and 10 treated. (B) The difference in number of successful trials for LWW compared to LWH (used as the gold standard) for each cell line, along with the associated risk type that these errors carry. Error types false positive (blue) and false negative (green) are shown.

reproducibility crisis affecting oncology drug trials, we cannot afford to ignore methods that greatly improve measurement accuracy.

**Contribution to the Field**

Our work, taken from a large dataset of real-world tumour measurements highlights a flaw in the most commonly used methodology for monitoring subcutaneous tumour volumes.

The assumption that tumour width is equal to height is incorrect, resulting in overestimation of tumour volume. Changes in tumour height are not recorded using the LWW volume calculation, and we have shown that height and width are not linearly correlated, so by ignoring height, information about treatment effect is lost. We also showed that 3D-TI measurement of height, length and width achieves greater accuracy when compared to a common calliper method. 3D-TI also has

advantages in comparison to existing imaging methods like ultrasound and MRI that offer improved accuracy in comparison to callipers including higher throughput, faster acquisition, and no reliance on reagents or anaesthesia.

Measuring treatment efficacy by monitoring subcutaneous tumour growth in vivo is a key step in preclinical studies. The oncology field has the lowest rate of drug development success, due to poor reproducibility and translation. More accurate results in preclinical stages will increase confidence in drug candidates chosen to progress to clinical trials.

## Acknowledgements

We thank Elisa Heyrman for her advice and edits to the manuscript.

## Author Contributions

Daniel Brough performed the analysis and wrote the Methods and Results. Hope Amos wrote the Introduction, Abstract and Discussion as well as providing scientific insight into the analysis process and cell line biology. Jake Murkin proof read, performed corrections and offered insight into the analysis process. Karl Turley framed the initial hypothesis, proof read and provided feedback throughout the analysis and writing of the results.

## Supplemental Material

Supplemental material for this article is available online.

## REFERENCES

- Euhus DM, Hudd C, LaRegina MC, Johnson FE. Tumor measurement in the nude mouse. *J Surg Oncol*. 1986;31:229-234.
- Dethlefsen LA, Prewitt JM, Mendelsohn ML. Analysis of tumor growth curves. *J Natl Cancer Inst*. 1968;40:389-405.
- Sápi J, Kovács L, Drexler DA, Kocsis P, Gajári D, Sápi Z. Tumor volume estimation and quasi-continuous administration for most effective bevacizumab therapy. *PLoS One*. 2015;10:e0142190.
- Lee JC, Lee HY, Moon CH, et al. Arsenic trioxide as a vascular disrupting agent: synergistic effect with irinotecan on tumor growth delay in a CT26 allograft model. *Transl Oncol*. 2013;6:83-91.
- Nguyen DX, Bos PD, Massagué J. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer*. 2009;9:274-284.
- Fontaine JJ, Marangoni E, Chateau-Joubert S, Servely JL. Pathology of patient-derived xenograft tumors. In: Uthamanthil R, Tinkey P, eds. *Patient Derived Tumor Xenograft Models*. Elsevier; 2017:135-148, Accessed June 28, 2022 <https://linkinghub.elsevier.com/retrieve/pii/B9780128040102000102>
- Tomayko MM, Reynolds CP. Determination of subcutaneous tumor size in athymic (nude) mice. *Cancer Chemother Pharmacol*. 1989;24:148-154.
- Feldman JP, Goldwasser R. A mathematical model for tumor volume evaluation using two-dimensions. *J Appl Quant Methods*. 2009;4:455-462.
- Jensen MM, Jørgensen JT, 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.
- Hall C, von Grabowiecki Y, Pearce SP, Dive C, Bagley S, Muller PAJ. iRFP (near-infrared fluorescent protein) imaging of subcutaneous and deep tissue tumours in mice highlights differences between imaging platforms. *Cancer Cell Int*. 2021;21:247.
- Ni J, Bongers A, Chamoli U, Bucci J, Graham P, Li Y. In vivo 3D MRI measurement of tumour volume in an orthotopic mouse model of prostate cancer. *Cancer Control*. 2019;26:1073274819846590.
- Ayers GD, McKinley ET, Zhao P, et al. Volume of preclinical xenograft tumors is more accurately assessed by ultrasound imaging than manual caliper measurements. *J Ultrasound Med*. 2010;29:891-901.
- Delgado San Martin J, Ehrhardt B, Paczkowski M, et al. An innovative non-invasive technique for subcutaneous tumour measurements. *PLoS One*. 2019;14:e0216690.
- Murkin JT, Amos HE, Brough DW, Turley KD. In silico modeling demonstrates that user variability during tumor measurement can affect in vivo therapeutic efficacy outcomes. *Cancer Inform*. 2022;21:11769351221139257.
- Brough DW, Murkin JT, Amos HE, Smith AI, Turley KD. Comparing variability in measurement of subcutaneous tumors in mice using 3D thermal imaging and Calipers. *Comp Med*. 2022;72:364-375.
- Thiel DD, Jorns J, Lohse CM, Chevillat JC, Thompson RH, Parker AS. Maximum tumor diameter is not an accurate predictor of renal cell carcinoma tumor volume. *Scand J Urol*. 2013;47:472-475.
- Workman P, Aboagye EO, Balkwill F, et al. Guidelines for the welfare and use of animals in cancer research. *Br J Cancer*. 2010;102:1555-1577.
- Wood S. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation version 1.8-40 from CRAN. Accessed July 18, 2022. <https://rdrr.io/cran/mgcv/>
- Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag; 2016. Accessed September 5, 2022. <https://ggplot2.tidyverse.org>
- Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Series B Stat Methodol*. 2011;73:3-36.
- Corbett TH, Griswold DP Jr, Roberts BJ, Peckham JC, Schabel FM Jr. Tumor induction relationships in development of transplantable cancers of the colon in mice for chemotherapy assays, with a note on carcinogen structure. *Cancer Res*. 1975;35:2434-2439.
- Gao ZG, Tian L, Hu J, Park IS, Bae YH. Prevention of metastasis in a 4T1 murine breast cancer model by doxorubicin carried by folate conjugated pH sensitive polymeric micelles. *J Control Release*. 2011;152:84-89.
- Yang S, Zhang JJ, Huang XY. Mouse models for tumor metastasis. In: Zheng Y, ed. *Rational Drug Design: Methods and Protocols*. Humana Press; 2012;221-228. doi:10.1007/978-1-62703-008-3\_17.
- Wu L, Wang W, Dai M, Li H, Chen C, Wang D. PPAR $\alpha$  ligand, AVE8134, and cyclooxygenase inhibitor therapy synergistically suppress lung cancer growth and metastasis. *BMC Cancer*. 2019;19:1166.
- Thalmann GN, Anezinis PE, Chang SM, et al. Androgen-independent cancer progression and bone metastasis in the LNCaP model of human prostate cancer. *Cancer Res*. 1994;54:2577-2581.
- Masuda J, Takayama E, Strober W, et al. Tumor growth limited to subcutaneous site vs tumor growth in pulmonary site exhibit differential effects on systemic immunities. *Oncol Rep*. 2017;38:449-455.
- Terracina KP, Aoyagi T, Huang WC, et al. Development of a metastatic murine colon cancer model. *J Surg Res*. 2015;199:106-114.
- Murphy KA, James BR, Wilber A, Griffith TS. A syngeneic mouse model of metastatic renal cell carcinoma for quantitative and longitudinal assessment of Preclinical Therapies. *J Vis Exp*. 2017;122:55080.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021. Accessed September 5, 2022. <https://www.R-project.org/>
- Huang L, Li Y, Du Y, et al. Mild photothermal therapy potentiates anti-PD-L1 treatment for immunologically cold tumors via an all-in-one and all-in-control strategy. *Nat Commun*. 2019;10:4871.
- Fox J, Weisberg S. *An R Companion to Applied Regression*. 3rd ed. SAGE; 2019. Accessed September 5, 2022. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Gao J, Li N, Dong Y, et al. miR-34a-5p suppresses colorectal cancer metastasis and predicts recurrence in patients with stage II/III colorectal cancer. *Oncogene*. 2015;34:4142-4152.
- Nagaraja GM, Othman M, Fox BP, et al. Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics. *Oncogene*. 2006;25:2328-2338.
- Gradiz R, Silva HC, Carvalho L, Botelho MF, Mota-Pinto A. MIA PaCa-2 and PANC-1 – pancreas ductal adenocarcinoma cell lines with neuroendocrine differentiation and somatostatin receptors. *Sci Rep*. 2016;6:1-14.
- Basu I, Locker J, Cassera MB, et al. Growth and metastases of human lung cancer are inhibited in mouse xenografts by a transition state analogue of 5'-methylthioadenosine phosphorylase. *J Biol Chem*. 2011;286:4902-4911.
- Fitzgerald JE, Byrd BK, Patil RA, et al. Heterogeneity of circulating tumor cell dissemination and lung metastases in a subcutaneous Lewis lung carcinoma model. *Biomed Opt Express*. 2020;11:3633-3647.
- Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. *Biom J*. 2008;50:346-363.
- Thalheimer RD, Merker VL, Ly KI, et al. Validating techniques for measurement of cutaneous neurofibromas: Recommendations for clinical trials. *Neurology*. 2021;97:S32-S41.