

Layman's summary

Rhabdomyosarcoma (RMS) is a soft tissue tumor that originates from muscle cells. For this reason, it can be found in various regions of the body: mostly in the head and neck, in the chest and abdomen, but also in the limbs. It is mostly diagnosed in children of ages 0-10, with a second peak around 15-19 years, with slightly more prevalence in male subjects. To image tumors, diffusion weighted imaging (DWI) is often used. It tracks the movement of molecules of water: fast moving groups of molecules are characterized by a higher apparent diffusion coefficient (ADC), which will appear brighter in the image. By scanning before and during treatment, it's possible to look at the change of ADC in the tumor due to chemotherapy. The broader aim of this research is to be able to distinguish between patients that respond well to treatment and those that don't, based on the change in ADC.

Since RMS is a rare tumor, many images from different hospitals are required in order to perform proper statistical analyses. Many of these hospitals use different acquisition parameters in order to obtain the scans, making it difficult to properly compare them. Since using different acquisition parameters leads to different ADC values in the image, we decided to study how much those parameters might vary within a database, and what differences in ADC values we have due to it. We collected images from a single center in the Netherlands (local database) and from centers across 5 countries in Europe (international database), from 27 and 73 patients. We found that the parameters in those images vary to a great extent. Both the patient characteristics as well as the acquisition detail. Furthermore, many of them did not follow the guidelines suggested for imaging this type of tumor.

The next step was to try to understand how much this variability leads to differences in ADC. We performed a statistical analysis, which did not show strong significant differences between images using different parameters. This is possibly caused by the small sample size to start with, that got reduced even more by missing information in the images. Moreover, some parameters can have the same effect on the ADC, so having such great ranges of parameters made the interpretation of results excessively complicated.

For this reason, we decided to perform an experiment. The shin of a single subject was scanned multiple times, varying one acquisition parameter at a time while keeping the others fixed. A statistical analysis was performed, which resulted in interesting findings: even small differences in acquisition parameters can change the mean ADC value of up to 17.7%, showing how carefully the acquisition parameter should be chosen. Needless to say, many more scans should be performed to have a stronger statistical analysis, possibly on tumors instead of healthy tissue.

RMS inevitably presents quite a large variability of patients' age and body part affected. Therefore, we should at least try to minimize as much as possible the variability caused by acquisition parameters. It is important, moreover, that the images are stored keeping as much information as possible for future studies.

The effect of diffusion MRI acquisition settings on Rhabdomyosarcoma images: a multicenter study

Author: Giulia De Donno

First examiner: Dr. Alexander Leemans

Second examiner: Dr. Alberto De Luca

Daily supervisor: Cyrano Chatziantoniou

Abstract

Rhabdomyosarcoma (RMS) is a soft tissue tumor that originates from the muscle cells. It is a rare tumor, therefore a multicenter study is needed to gather a sufficiently large amount of images. In this study, we compared two datasets of diffusion weighted images (DWI) coming from a single center in the Netherlands and from centers across Europe. The databases, respectively of 51 and 126 images from 25 and 73 patients. Acquisition parameters and patient information were analyzed to quantify inter and intra center variability and the effect of those on mean apparent diffusion coefficient (ADC). From the two datasets, high variability has been observed in acquisition parameters and patients' information. No particular links to the ADC values was found, possibly due to the large amount of data missing from some of the images, especially b-values. In the second part of the study, a DWI scan on a healthy subject was performed using echo time, voxel dimension and b-values most commonly found in the datasets. Varying one parameter at a time and keeping the other fixed, a total of 10 images were acquired. The aim was to isolate the effect of each acquisition parameter on mean ADC and differences from 1.0% up to 17.7% were found in mean ADC. These findings show how important it is to adhere to a strict set of acquisition parameters, in order to minimize as much as possible their variability and influence on ADC values.

Introduction

Rhabdomyosarcoma (RMS) accounts for around 50% of soft tissue sarcomas in pediatric patients and 3-5% of all pediatric malignancies.^{1,2} It originates from primitive mesenchymal stem cells, which can differentiate in muscle cells. For this reason, these cells can originate RMS in a variety of anatomical sites throughout the body, especially head and neck, chest, abdomen, genitourinary tract, and extremities.^{3,4} As shown in [Table A.1](#) in the appendix, RMS manifests around 44% in the head and neck region (parameningeal, orbital, head and neck), 34% in the trunk (bladder, prostate, trunk, retroperitoneal, perineal/perianal, genitourinary), 14% in the extremities and for a 12% in other parts of the body.^{1,2,5,6} Around 70% of cases are diagnosed within the first 10 years of life, with a second peak around 15-19 years. The prevalence is slightly higher in male subjects (~60%) compared to females (~40%), but sex or race have had no influence on survival.^{1,2,5,6}

Treatment for RMS is now based on a multimodal approach, based on cycles of multidrug chemotherapy, surgery and/or radiotherapy, followed by maintenance chemotherapy. Both treatment plan and imaging acquisition parameters are established by the European rhabdomyosarcoma imaging guideline.⁷ The volume of the tumor can be used as a method to determine prognosis and changes in treatment course in case of progression of the disease, but it cannot be used as marker for survival. In fact, currently, there

are no early biomarkers that can help identify a positive or negative response to the chemotherapy regime, and survival is the only valid endpoint. ² Future studies will try to evaluate if changes in apparent diffusion coefficient (ADC) might be a good predictor.

Rhabdomyosarcoma is a rare tumor, diagnosed in around 15-30 patients in the Netherlands each year. As a large database of diffusion MR images is required, a multicenter approach will be necessary. ² When working with multicenter data, it's important to be aware of the bias that comes from differences in hardware and software among different MRI scanners.⁸ For example, variabilities might arise from the use of different receiver and transmitter coils, or variations in magnetic field inhomogeneities. Furthermore, acquisition parameters such as voxel size, number of gradient directions, echo time, and the b-values used can all affect the measured ADC. ⁹ As rhabdomyosarcoma pediatric patients range from birth to late teens, additional variability is introduced not only in the wide range of body regions, but also in dimension of the body itself.

This large range of factors might lead to a substantial difference in the ADC. It is of interest, therefore, how well images that were acquired with different scanners and with different scanning parameters can be compared. In this study, the aim is to determine the amount of inter and intra-center variability in multicenter RMS data. We furthermore investigate the effect of the variability on ADC by scanning a healthy volunteer with the most common imaging parameters found in the analysis.

Material and Methods

The study was divided in two parts: in the first part of the study, two datasets are analyzed, with focus on intra and inter-center variability of acquisition parameters and patients' personal information. In the second an experiment is conducted to isolate the effect of acquisition parameters on ADC values.

Retrospective study

The dataset from a single center (local) amounted to 51 images, coming mostly from the University Medical Center Utrecht (UMC, Utrecht, NL), and the Princess Máxima Center for Pediatric Oncology (PMC, Utrecht, NL). It had had a total of 25 patients, with 25 images at diagnosis and 26 at follow up. Data from the header was collected from each of them. A subset of 17 ADC maps with the same type of segmentation was further analyzed. The second dataset (international) had 126 images that come from centers in Italy (55), France (21), Spain (8), Norway (15), and other centers in the Netherlands (27). Of the 126 images, 73 images were made at diagnosis and 53 at follow up, for a total of 73 patients.

From each patient, DWI, ADC maps or diffusion tensor images were received with anonymized personal data about the patient and acquisition parameters stored in the header. Amid many parameters, the main personal information obtained from each scan were: age, sex and body part scanned. Date of diagnosis, type of image (ADC or DWI), information about the 3 dimensions of the voxel, echo time (TE) and repetition time (TR), b-values and field strength were some of the key information collected to describe the scan. The anisotropy ratio ([Eq. 1](#)), is a coefficient that can vary from 0 to 1 that links slice thickness and pixel spacing of the voxel and it was calculated for each scan. An anisotropy ratio of 1 indicates an isotropic (cubic) voxel.

$$\text{Anisotropy Ratio} = 1 - \frac{\text{Pixel Spacing [mm]}}{\text{Slice Thickness [mm]}} \quad (\text{Eq. 1})$$

ADC maps were calculated based on DWI or DTI images when not already present. In each image, the tumor was segmented with consensus on a single axial slice by radiologists (SH and RR) with 10 and 18 years' experience. The ADC values from the region of interest (ROI) were then analyzed. Mean ADC and ADC at percentile 0, 25, 50, 75 and 100 were extracted from the ROI. A normality test (Shapiro-Wilk) was performed on the mean ADC vector. Then, relationships between mean ADC and other parameters in the database were investigated by one-way ANOVA test. For each parameter, mean ADC above and below a threshold was calculated and compared using a t-test. The median value was mostly used as a threshold, however parameters with a very low number of unique values were analyzed over 3 categories. The difference was considered significant only for p-values < 0.05.

All the data was also analyzed to quantify inter and intra center variability.

In-vivo experiment

In order to look at the impact of varying scan parameters, we scanned a healthy volunteer age (27 years), using the most frequently occurring values of b-values, echo time, and voxel size as range of parameters. These scans were performed on the shin of a healthy subject on a single scanner, varying one parameter at a time, and keeping the others fixed. A total of 10 diffusion weighted images and a structural image T2 image were acquired using a Philips Achieva 3T scanner, from the University Medical Center Utrecht (UMC, Utrecht, The Netherlands).

As shown in [Table M.1](#), three groups of images were acquired: I) Scans where the echo time ranged from 50 ms to 110 ms in 4 steps of 20 ms (4 images). II) Scans where the voxel size was ranged from a ratio of pixel spacing and slice thickness of 2:2, 2:4, 2:6, 2:8, 1:8 (5 images, [Figure M.1](#)). III) A single scan with a larger number of b-values, ranging from 0 to 1200 s/mm² in 7 steps. A total of 5 ADC maps were calculated using different combinations of b-values and compared. The combinations analyzed are described in [Table M.1](#), and they reflect some of the most used combinations of b-values found in the local and international datasets and the guidelines.

The base parameters were: echo time of 50 ms, voxel dimensions [2 x 2 x 4] mm, b-values [0, 50, 500, 1000] s/mm², 22 slices, 64 rows and columns, field strength of 3T.

The ADC values were obtained by delineating the outline of the muscle on a single axial slice in the middle of the scanned volume. This was performed by a trained researcher. The changes in the ADC were then analyzed using the Shapiro-Wilk test for normality, one-way ANOVA to test for statistical significance in ADC values in images where a parameter varied and a Tukey's HSD (honestly significant difference) to test for pairwise differences (p-value < 0.05). The aim is to try to quantify how changes in the acquisition parameters influence the mean ADC of the tissue.

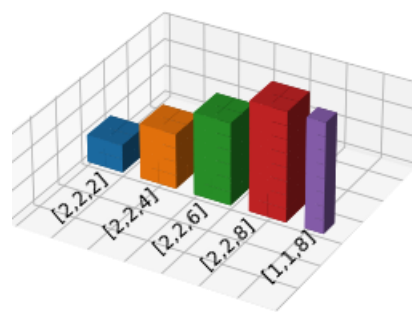


Figure M.1 Visual representation of the voxels used in the in-vivo experiment (part II, voxel dimension varied). The values expressed are in mm [x,y,z].

Table M.1 Acquisition parameter of scan on the shin of a healthy subject. A number of 10 images and 14 ADC maps were analyzed in the second part of the study (in-vivo experiment).

		Voxel size [mm]	Echo Time [ms]	B-value [s/mm ²]
Group 1: TE varies	Scan_TE_50	2 x 2 x 4	50	0, 50, 500, 1000
	Scan_TE_70	2 x 2 x 4	70	0, 50, 500, 1000
	Scan_TE_90	2 x 2 x 4	90	0, 50, 500, 1000
	Scan_TE_110	2 x 2 x 4	110	0, 50, 500, 1000
Group 2: Voxel dimension varies	Scan_VOX_222	2 x 2 x 2	50	0, 50, 500, 1000
	Scan_VOX_224	2 x 2 x 4	50	0, 50, 500, 1000
	Scan_VOX_226	2 x 2 x 6	50	0, 50, 500, 1000
	Scan_VOX_228	2 x 2 x 8	50	0, 50, 500, 1000
	Scan_VOX_118	1 x 1 x 8	50	0, 50, 500, 1000
Group 3: B-value varies (1 scan*, 5 ADC maps calculated)	8-1200 *	2 x 2 x 4	50	0, 171, 343, 514, 686, 857, 1028, 1200
	2-1200	2 x 2 x 4	50	0, 1200
	2-1028	2 x 2 x 4	50	0, 1028
	4-1028	2 x 2 x 4	50	0, 171, 514, 1028
	5-1028	2 x 2 x 4	50	0, 171, 514, 857, 1028

Results

Retrospective study

In this study, two datasets are compared. Information from 51 images from 25 patients and 126 images from 73 patients are analyzed from the local and international database. Of the 51 images of the local database, 40 are pre-calculated ADC maps, 8 contained both ADC and diffusion weighted images (DWI), only 3 contained diffusion weighted images. Of those, ADC values from 17 of those are analyzed. Of the 126 images of the international database: 118 images are pre-calculated ADC maps and the others were just DWI or DTI; 99 of those were used in the ADC calculations.

Patient information:

For the local database, the mean age is 8 ± 5.3 years. For the international database, 26% of the images did not have this information. From the data available, mean age calculated is 6.5 ± 5.3 years ([Figure A.1](#)). In the local database, 25% of the patients were female, 75% male. In the international database 31% of the patients were female, 58% male, and for 11% of the patients, the information was unavailable ([Figure A.2](#)).

The mean time between two scans is 64 ± 13 days in the local database. In the international one, the images that had both diagnosis and follow up images have a mean difference of 74 ± 36 days. The year of acquisition ranged from 2015 to 2020 in the local, and from 2010 to 2021 for the international. In the local database, 47% of the images were scanned from the trunk (abdomen, chest, pelvis, and genitourinary region), 20% from the extremities (arms, hand, legs and feet), 33% from head and neck. In the international dataset, in 40% of cases the information had to be visually extrapolated from the image itself (it was not present in the header); 27% of the images are from the trunk, 4% from the extremities and 69% from head and neck region ([Figure R.1](#))

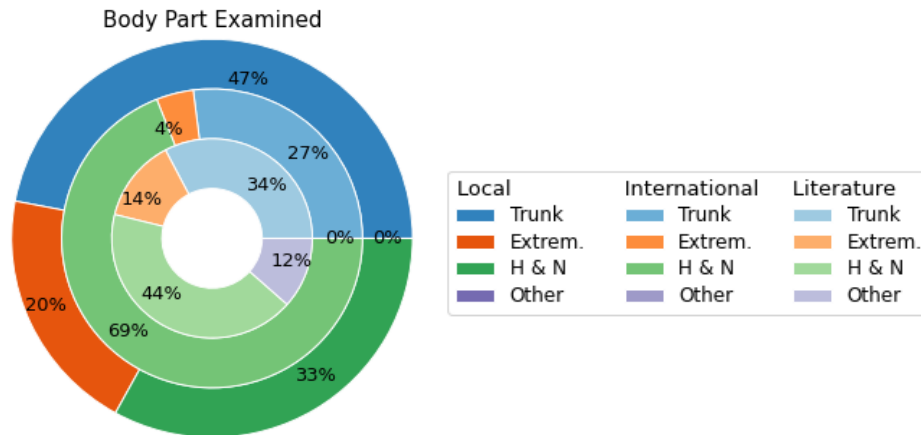


Figure R.1 Representation of the body part examined in the local and international database compared with the literature.

Scanning information:

For both datasets, the echo time had similar ranges, both varying from a minimum value of around 50 ms to around 100-110 ms. The mean is for the local 74 ± 9.5 ms while for the international is 73 ± 24 ms (Figure A.3). The repetition time has a wider range of values: from the local dataset the mean is 3740 ± 2650 ms, with a minimum value of 1350 ms and a maximum value of 12800 ms. Similarly, the international database has a mean value of 5130 ± 1920 ms. The range is between 1360 ms and 12780 ms (Figure A.4).

Regarding the voxel size, the following parameters will be addressed: pixel spacing (the width and depth of the voxel, described from now on as a single variable, since the pixel is a square), the slice thickness (the height) and the anisotropy ratio, which is a coefficient that links those parameters together.

Both the local database and the international database have similar ranges for pixel spacing: the mean value was 1.20 ± 0.41 mm and 1.35 ± 0.45 mm. However, as presented in Figure R.2, the internal distribution of the sizes is different. In the two images, it's shown how the different body parts are linked to the pixel size.

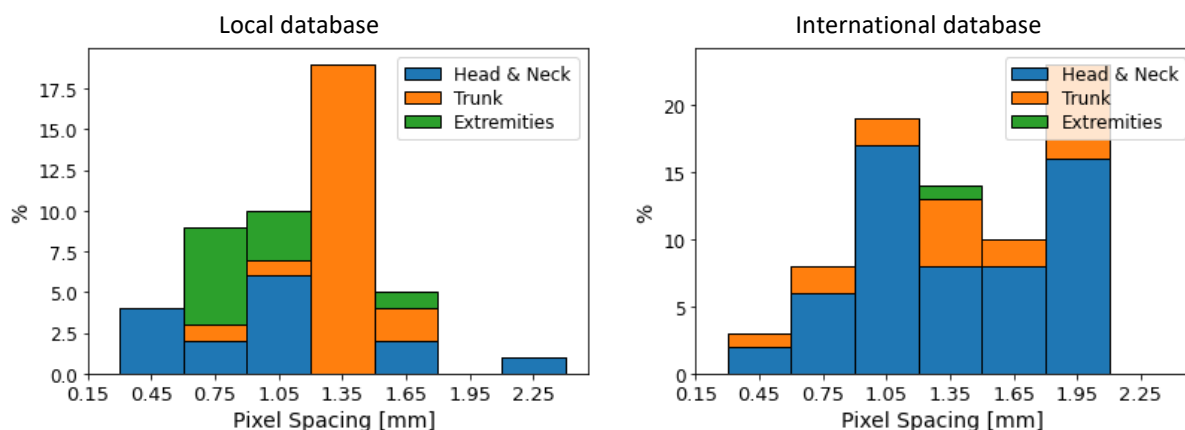


Figure R.2 Representation of the distribution of pixel spacing for the different body parts.

As we can see in Figure A.5, the local dataset has a slice thickness that varies from 3 to 7 mm, whereas the international dataset varies from 2 to 7 mm. The most common value is 5 mm in both cases (45% and

39% of the total). In the local dataset, the largest slice thickness is mostly used for the extremities and the smallest for head and neck. In the international dataset we can see that the abdomen is scanned with a slice thickness varying from 3 to 7 mm, showing the biggest range of all the body parts (more information on [Figure A.7](#)).

As it can be seen in the [Figure A.6](#) for each body part the anisotropy can vary quite extensively in the two datasets, especially in images from the head and neck. In the appendix, in [Figure A.8](#) it is presented how the voxel with the highest and lowest anisotropy ratio looks for each body part. [Figure R.3](#), indicates the different combination of pixel spacing and slice thickness used in each dataset, highlighting the resulting anisotropy ratio.

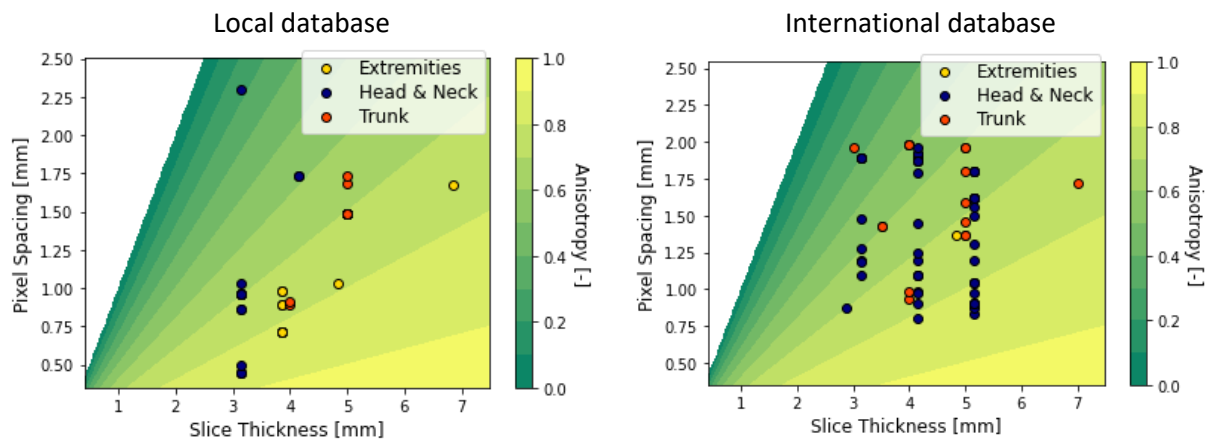


Figure R.3 Representation of pixel spacing and slice thickness for both databases. In the background, the anisotropy represented in color.

The b-value was missing in 38 images out of 51 (75%) in the local dataset, in the international in 97 out of 120 (81%). As can be seen in [Figure R.4](#) the b-values ranged from 50 to 1000 and the combination 0-1000 s/mm² is the most common. Most of the images had only 2 b-values, but some images also had up to 5 shells (e.g. 0, 50, 400, 800, 1000 s/mm²).

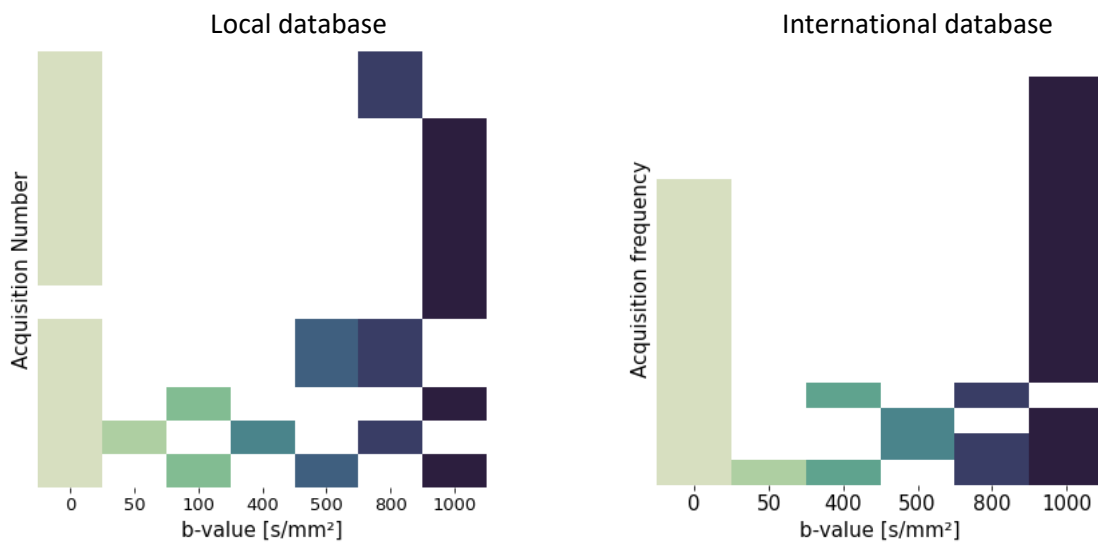


Figure R.4 represents the combinations of b-values found in the databases.

A Shapiro-Wilk test was used to determine if the mean ADC of the images was modeled by normal distribution. The local dataset, with only 17 values, proved to be normally distributed (p-value = 0.1145), as well as the international database, with 99 images (p-value= 0.1029).

A one-way ANOVA test was performed on the mean ADC vectors, using median as a threshold or grouping values in 3 categories. The test showed a significant difference in mean ADC for images taken at diagnosis and follow up of 41% and 30% for local and international dataset (median 50 days, local: p-value 0.017, international: <0.001) (Figure R.5). In the local dataset, the anisotropy also showed significant differences for values above and below the median value 0.7 (p-value: 0.037, 53%), whereas the same parameter did not show the same trend in the international database (median 0.68, p-value 0.346, 6%). No other significant differences in ADC values were found (p-value > 0.05). It is to be noted how for certain parameters almost half of their values were missing, especially for the b-value, present in only 4 images over 17 and 9 images out of 99 (Table A.2).

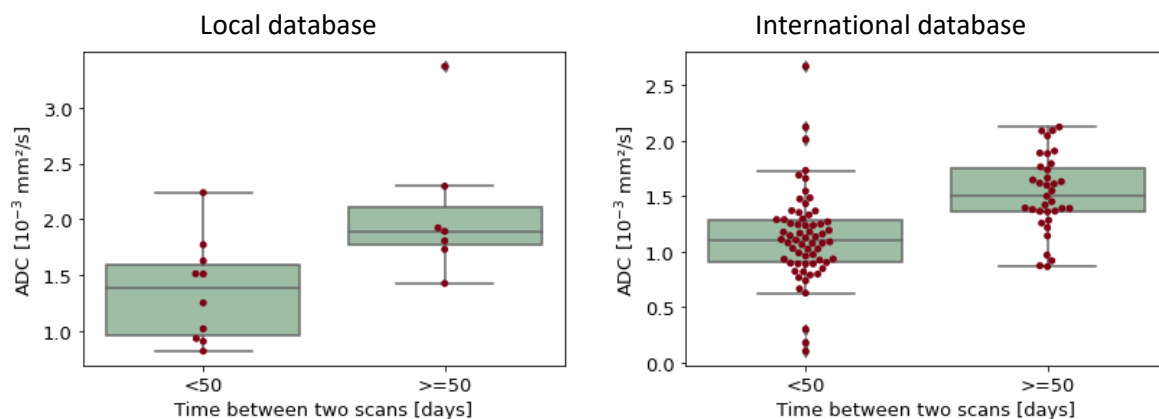


Figure R.5 Box plot of the ADC values above and under the threshold (median). The difference between the two groups is significant, with p-values 0.017 for the local database and for the international: <0.001.

In-vivo experiment

A second experiment was performed scanning a healthy subject varying one of three acquisition parameter at the time, keeping the others fixed (Table M.1). The three groups of images were analyzed separately. The ADC values from all of them proved to be distributed normally at the Shapiro-Wilk test. A one-way ANOVA test was conducted from ADC values of each group, to check for significant differences linked to the change in parameters.

In group 1, the echo time was varied and a significant difference was found within the 4 images (p-value <0.001), therefore a pairwise comparison of the images was performed. Differences in the ADC values from these images varied from a minimum of 2.1% (TE of 70 ms and 90 ms) to a maximum of 17.7% (TE 50 and 110 ms) (Figure R.6.a).

In group 2, the voxel dimension changed. The same tests showed significant differences in ADC between the 5 images (p-value <0.001) from 1.1% ([2,2,4] to [1,1,8] mm) to 17.7% ([2,2,2] to [2,2,8] mm). It is to be noted how in the first couple, the anisotropy ratio varies from 0.5 to 0.875 and the volume halves (from 16 to 8 mm³), whereas in the second from 0 to 0.75 and the volume increases fourfold (from 8 to 32 mm³) (Figure R.6.b, Figure M.1).

For the third group, images with different combinations of b-values were calculated and compared (p -value < 0.05). The smallest significant difference was 2.8% and it was between '2-1028' and '2-1200' (b-values [0, 1028] and [0, 1200] s/mm²), where the same amount of b-shells are used, but the maximum b-values varies. The largest difference found was 8.8%, between '2-1028' and '4-1028' (b-values [0, 1028] and [0, 171, 514, 1028] s/mm²) where the maximum b-values in the same, but double the amount of b-values are used to create the ADC map (Figure R.6.c, Table M.1).

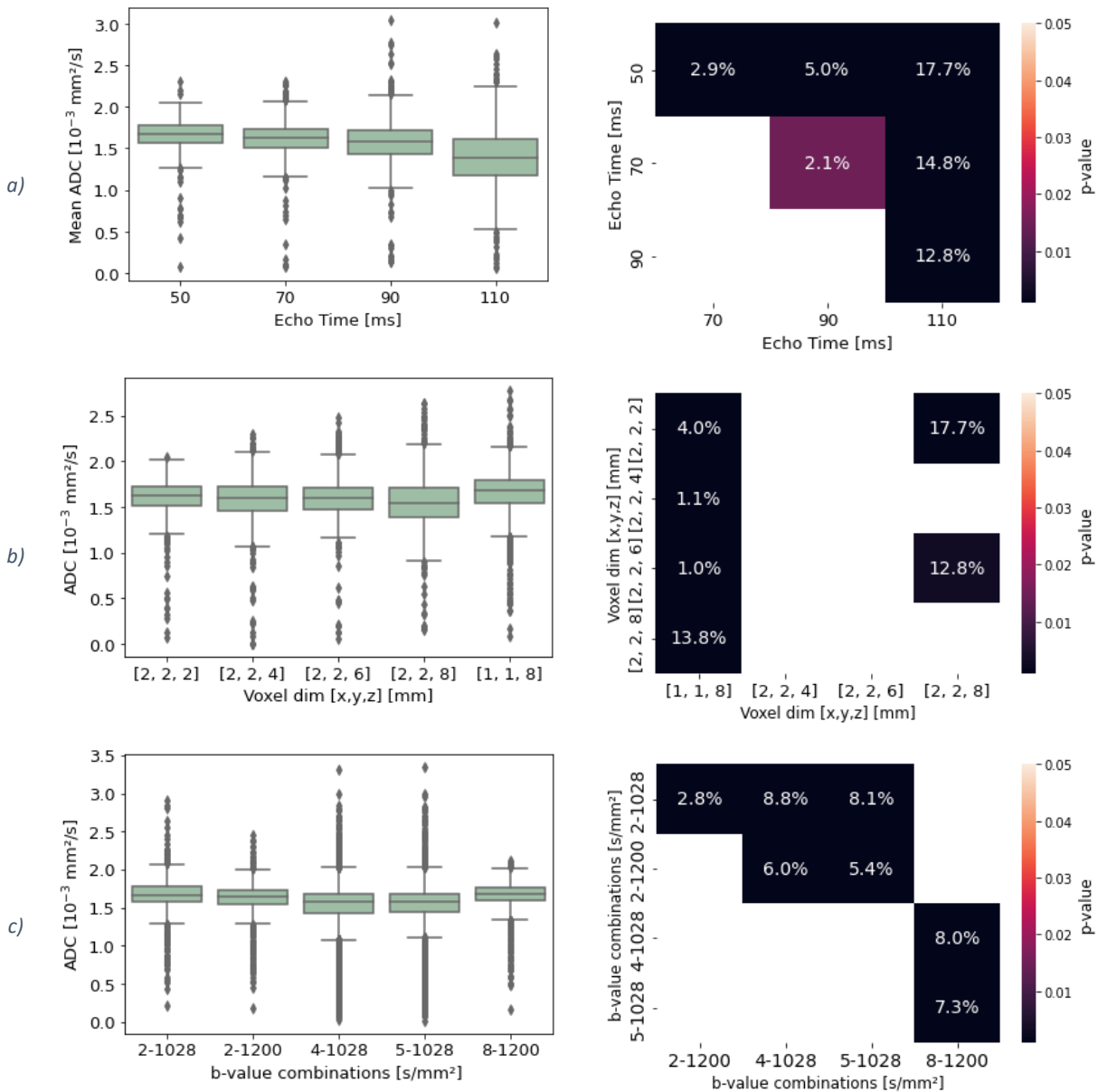


Figure R.6 a,b,c Result of the statistical analysis of the scan of the shin. Each group (different echo time, voxel dimension and b-value combination), was analyzed separately. In each group, a one-way ANOVA test was used to confirm significant difference between each sub-category, then a pairwise test was performed. For each group we can see on the left the box-plot of the ADC values of the images. On the right, a heatmap containing only pairs with significant differences in ADC values (p -value < 0.05) and the respective percentage difference between mean ADC. The color indicates the respective p -value for each pair. We can see that when the different is significant, most p -values are below 0.01

Table D.1 Data from the EpSSG Rhabdomyosarcoma Imaging guideline (May, 2021), valid for Philips, Siemens and GE scanners.

Field strength [T]	Body Part	B-value [s/mm ²]	FOV [mm]	Slices	Slice Thickness [mm]	Voxel Size [mm]	TE [ms]	TR [ms]
3.0	Head and Neck	0, 100, 500, 1000	230 x 196	41	3.0	2.0 x 2.0	Shortest 75	4000-5000
3.0	Extremities	0, 100, 500, 1000	230 x 230	50	4.0	2.5 x 2.5	Shortest 54	4000-5000
1.5	Head and Neck	0, 100, 500, 1000	200 x 200	41	4.0	2.81 x 2.81	Shortest 71	4000-5000
1.5	Chest and Abdomen (small child, with anesthesia)	0, 100, 500, 1000	380 x 332	26	5.0	2.81 x 3.5	Shortest 78	4000-5000
1.5	Chest and Abdomen (large child, without anesthesia)	0, 100, 500, 1000	380 x 380	45	5.0	2.81 x 3.49	Shortest 78	4000-5000
1.5	Extremities	0, 100, 500, 1000	200 x 200	40	4.0	2.5 x 2.5	72	4000-5000

Discussion

Retrospective study

In this study, a dataset from a single center (local) and a dataset from multiple centers (international) were analyzed, with particular focus on intra and inter-center variability of parameters.

As we have seen, there is a large heterogeneity of results coming from the two datasets. The patients range from newborns to late teen, which results in large differences in body size and therefore scan volume. Additionally, the body part that was scanned can further affect the results. Scanning extremities, for example, can bring some type of artifacts that are rarer in other parts of the body. For example, motion artifacts caused by breathing are more common in the trunk, whereas Gibbs artifacts are more common in brain and spine imaging.¹⁰ The EpSSG Rhabdomyosarcoma Imaging guideline (May, 2021) indicates which acquisition parameters should be used for the different body parts (head and neck, chest and abdomen, extremities), scanner manufacturer, and field strength (1.5T or 3.0T) used (Table D.1).

For DWI with ADC maps, a minimum of 4 b-values (including b=0) is recommended, with the other in ranges between 0 and 200, 200 and 800, 800 and 1200 s/mm². In the database, we saw images with b-

values ranging from 0 to 1000 s/mm². The most common combination was 0 -1000 s/mm², which is not stated in the guidelines.

For this study, one of the main problems encountered was the amount of data missing from the headers, in particular the b-values. The b-value is one of the most important parameter in diffusion MRI and it is indispensable to calculate the ADC map from a DWI. From the different type of images collected for this study, it is noticeable how the ADC maps were the ones where the b-values were discarded the most. The ADC maps, however, not only highlight diffusion visually as intensity values, but also contain quantitative data, that depends on the b-value used. In order to be able to compare information from different ADC maps, it is crucial to take in account this parameter. It is also important to consider how there are different degrees of approximation in calculating the ADC maps from DWI that can bring to slightly different results. It would be ideal to receive the DWI and the b-values and then being able to calculate the ADC map with the same formula for all the images.

Another parameter that will influence the contrast in the image is the echo time. A longer echo time will lead to a less intense signal, leading to a lower signal-to-noise ratio. In our study, the TE varied from around 50 ms to around 100 ms ([Figure A.3](#)) and their mean value is consistent with the guidelines.

As we can see in the guidelines, there are clear indications regarding the dimension of the voxel. The slice thickness indications were followed by most of the images in the local database, whereas in the international diverged more. The instructions for the pixel size (indicating the side length of the square pixel), on the other hand, were not followed in most of both datasets.

The pixel size in the databases was usually smaller compared to the one indicated in the guidelines. This leads to a higher in-plane resolution, however an unintended consequence is that the voxel becomes more anisotropic ([Figure A.8](#), [Figure R.3](#)). Having a voxel that has a dimension predominantly longer than the others might lead to different ADC values depending from the tissue structure and the voxel orientation. This is be predominant in tissues that have more anisotropic microstructure, for example in the brain.

Since an in-plane resolution as high as possible is favorable, a solution to this problem could be to reduce the slice thickness. This would reshape the voxel into a more cubic shape, reducing the anisotropy ratio. However, it would require a larger number of slices to image the same portion of the body with a smaller slice thickness, leading to a longer total time of scanning, being more expensive both time and money-wise. Furthermore, it would need more time for a physician to analyze the image and longer segmentation times. A more in depth study on the effects of the variation of anisotropy ratio in rhabdomyosarcoma could lead to interesting finds, which could reshape future guidelines.

In order to further understand if the use of different parameters could influence ADC values in images in the databases, some statistical tests were performed. The analysis was performed separately on the two datasets and it focused on one parameter at the time. The parameters analyzed comprehended acquisition parameters of the scan and personal information of the patient.

For each parameter, all of its values were divided into two or more subgroups and the respective ADC values were analyzed. A one-way ANOVA test was performed on ADC values from images of each group, to see if there is a significant difference between them.

Of all the 19 parameters analyzed in both datasets, most of them did not show a significant difference between the groups. However, it in both datasets, ADC values in the segmented tumor showed differences with statistical significance from diagnosis to follow-up. In both cases, the mean ADC increased of up to 41%. This is compatible with what we know from the literature: anticancer therapies induce cell death or damage, leading to increase of mobility of water in the tissue, therefore translating

in a higher measured ADC.¹¹ It is worth mentioning that in both datasets images at diagnosis represent 60-64% of the totality of the images analyzed.

In the local database (but not in the international one), differences in anisotropy ratio showed statistical significance. Values above 0.7 showed an increase of mean ADC of 53%. Interestingly, differences in volume showed no statistical significance, suggesting that it might be the different shape of the voxel that leads to different ADC values, not the difference in volume. Images with lower anisotropy (therefore more cubic in shape), showed overall lower ADC values. It is unexpected to have found this relationship only in the local database, and it might be linked to the fact that in the local database most of the images were from the trunk, whereas in the international came from the head and neck region, therefore they might have different properties. However, it is worth remembering that even if it this difference in ADC values has a statistical significance, it came from a small sample size, having a low statistical power.

It is to be noted how the number of b-values or shells used seemed not to influence the ADC values. This is an unusual behavior, which could be simply explained by the small amount of data available. For the local database, only 4 of the 17 images had information about the b-value, for the international only 9 over 99 images. This was probably not enough to make a statistically significant difference.

A similar limitation can be imagined for other parameters: a larger sample of images would have possibly shown more significant results.

In-vivo experiment

Due to the variability of data within each database, it was decided to try to isolate the effect on ADC from acquisition parameters with an experiment. The scan was performed on a single subject, on a single scanner, with one operator to segment the image, and calculating the ADC maps with the same software, to limit variability. Then one parameter at the time was varied keeping the other fixed.

In this way, the interdependency between the acquisition parameters is limited as much as possible. The ranges of acquisition parameters in those scans were chosen using the most common ranges of those variables in the datasets. An interesting point to prove was to see if we did not encounter significant differences in the local and international datasets because there were not any or because of the links between variables, masking some of their effects.

From this experiment, we overall found how even small differences between acquisition parameters can lead to significant differences in ADC values. In the first group, where the echo time varied, we could see differences in ADC values up to 17.7%, and even how differences in TE of 20 ms can change the ADC between 2.1% (70-90 ms) and 12.8% (90-110 ms) ([Figure R.6.a](#)). In the international and local datasets, the ranges of TE were wide, and it is possible to imagine how those ADC valued could have been biased just by the TE alone.

The voxel dimension as well led to interesting results. Keeping the pixel spacing at 2 mm and varying the slice thickness from 2 (isotropic) up to 6 mm did not lead to differences with statistical significance, however a rise from [2,2,6] and [2,2,8] mm led to a difference of 12.8%, with a total difference of 17.7% between [2,2,2] and [2,2,8] mm ([Figure R.6.b](#)). Of course, with the increment of the slice thickness, the volume of the voxel increased, as well as the anisotropy ratio. It seems that for values with anisotropy ratio under 0.6, the volume difference does not play such as an important part as it does above that threshold. We can see that as well between voxel [2,2,8] and [1,1,8] mm, with a 13% difference. In this case, the anisotropy just slightly increased (from 0.75 to 0.88), but the volume decreased by a factor of 4. It is interesting to compare this result with the one from the local database, where it was only the anisotropy ratio that lead to a difference of statistical significance.

It is therefore important to keep in mind how the volume of the voxel and its anisotropy could lead to changes in ADC values. It would be beneficial to update the guidelines, giving voxel dimension that would be useful to the clinician for diagnostic purposes, but also have a low anisotropy ratio, from an imaging point of view.

The different ADC maps that were calculated with the use of different b-values did not show such dramatic differences. It is interesting to note however how keeping the same maximum b-value but using 2 shells instead of 4 could mean a difference of 8.8%, and using 5 of 8.1% in the ADC values ([Figure R.6.c](#)). This underlines how important it is that the images should be scanned following the guidelines, which state ranges of values to be used.

As emerged through the retrospective study, one of the main issues encountered was the quantity of missing data. Especially for the local dataset, already smaller in size, this brought some difficulties in the statistical analysis. The same can be said for the second part of the study, the experiment. Only one single subject was analyzed, for only one session. This, of course, limits the statistical power of the experiment. Ideally, more healthy patients should be imaged on different scanners, possibly more than one time. Once a pattern is established, it would be interesting to scan some patients, to try to isolate the effects on ADC on the tumor itself.

This study leads to the final consideration of how important it would be to update the guidelines, giving a smaller range of value for each parameter, and to inform the clinicians how to deviate from those guidelines could lead to different results. Since it might be difficult to follow the guidelines in some cases, it is also worth reminding how valuable would be for the clinicians to store as much information as it is possible in the image, especially b-value and echo time, and possibly to store the DWI as well as the ADC maps.

Knowing what type of bias on the ADC values is due to acquisition parameters, it is crucial for a multicenter study and future investigations on ADC values, with the final aim being finding a reliable early biomarker. Being able to understand if a therapy is working in less time would allow for more new treatments to be tested, and to change the course of therapy if a treatment is not working. Moreover, this could help tailor the right therapy to the patient, being less strong if the patient responds well (to decrease side effects), or adjusting to a new therapy in case the one used is failing. This could considerably improve the quality of the results of treatments and the quality of life for the patients.

Conclusion

Due to the rarity of Rhabdomyosarcoma, a multicenter study is needed in order to get a larger amount of images. In this study, a single and a multi-center datasets were analyzed. It emerged that even within a single center, there is high variability in acquisition parameters, which gets even greater when the multicenter database is analyzed. Since the objective is to start a multicenter study, it is crucial to understand if the inevitable variability caused by patient information and the one caused by scanner and acquisition parameters can influence the ADC values. From the investigation on the datasets, a clear link between parameters and ADC values does not appear, if not from the time between diagnosis and follow up. Other parameters, expected to have a relationship with ADC values (e.g. echo time or b-value), did not seem to have an effect on it. This might be explained by the limited amount of data to analyze or the interdependence between parameters.

At this end, it was decided to scan a healthy subject. From this experiment, it emerged that even small changes within echo time, voxel dimension or b-value can bring considerable significant differences between images. Consequently, it would be of the outmost importance to try to minimize the differences in acquisition parameters between scans as much as possible, following the guidelines in already in place. Another important point to stress in the necessity to store all information available of the image and DWI images, indispensable to enable a multicenter study.

Appendix

Table A.1 Prevalence of body part scanned in literature and local and international database in percentage. (H & N = Head and Neck)

		Sultan et al. (2009) ¹	Gaal et al. (2012) ⁵	Quaglia et al. (1994) ⁶	van Ewijk et al. (2021) ²					Mean Value % Literature	Local database	International database
					Burke et al. (2007)	Dantonello et al. (2015)	Ferrari et al. (2009)	Rosenberg et al. (2014)	Vaarwerk et al. (2017)			
H & N	Parameningeal	10	23		40	37	24	46	31	44	33	69
	Orbit	5	4		11	14		4	14			
	Head and Neck	14	8	26	4	6	17	2	10			
Trunk	Bladder and prostate	6	9		13	17	6		15	34	47	27
	Trunk	22	4	13			16	12				
	Retroperitoneal	3	4									
	Perineal and perianal		3									
	Genitourinary	15	23	35	7	5	25	13	6			
	Extremities	19	13	27	9	3	12	15	11	14	20	4
	Other	7	9		18		9	13	15	12		

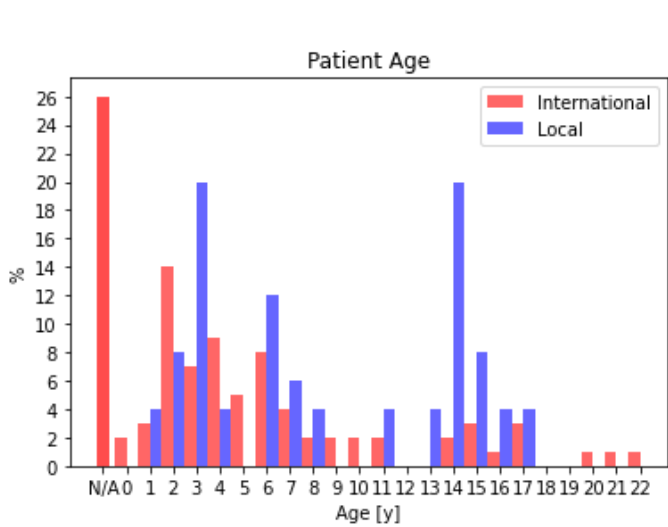


Figure A.1 Histogram of age of patients for local and international database. The local database has a behavior similar to the literature (bimodal, with peaks around 3 and 15). The mean age is 8.0 ± 5.3 years. In the international, 26% of the information is missing; patients are mostly above 10 years old, with a lower mean age (6.5 ± 5.3 years).

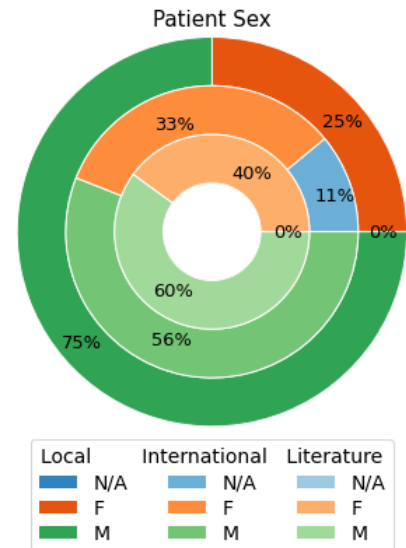


Figure A.2 Pie chart indicating the prevalence of female or male patients in the local and international database, compared with the literature. Of the international database, 11% of information were missing.

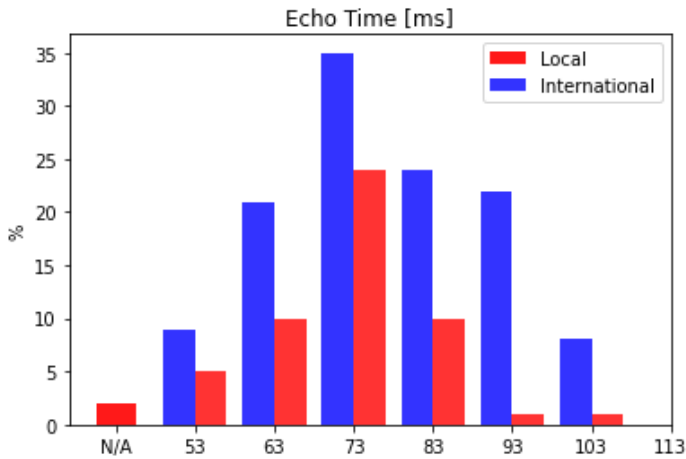


Figure A.3 Percentage of echo time values in local and international database expressed in ms. The mean is for the local database is 74 ± 9.5 ms, for the international is 73 ± 24 ms.

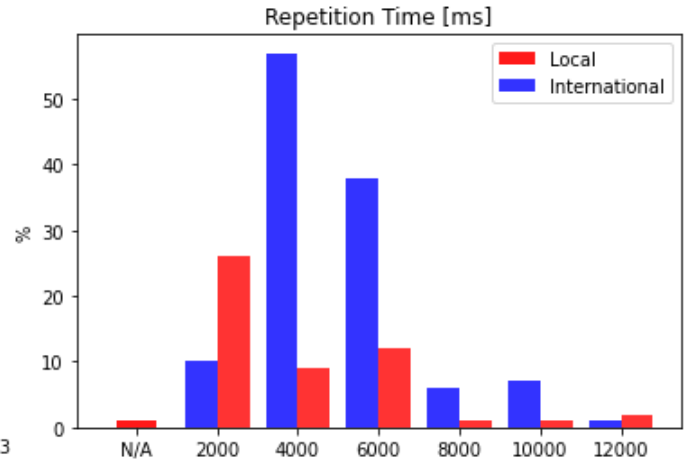


Figure A.4 Percentage of repetition time values in local and international database expressed in ms. The mean value is 3740 ± 2650 ms for the local database and 5130 ± 1920 ms for the international one.

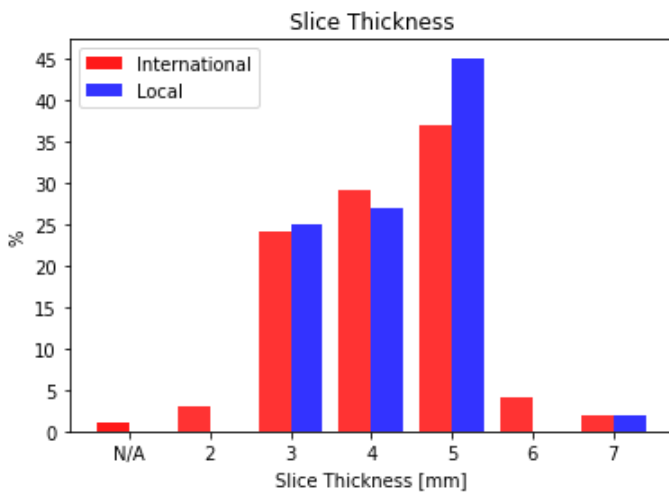


Figure A.5 Percentage of slice thickness values in local and international database expressed in mm. The international database, even if mostly has scans of head and neck, has the highest spread of values.

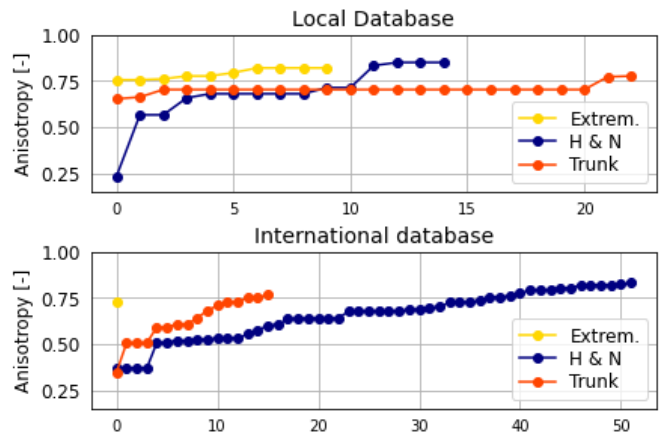


Figure A.6 Plot of the anisotropy ratio for each category of body type.

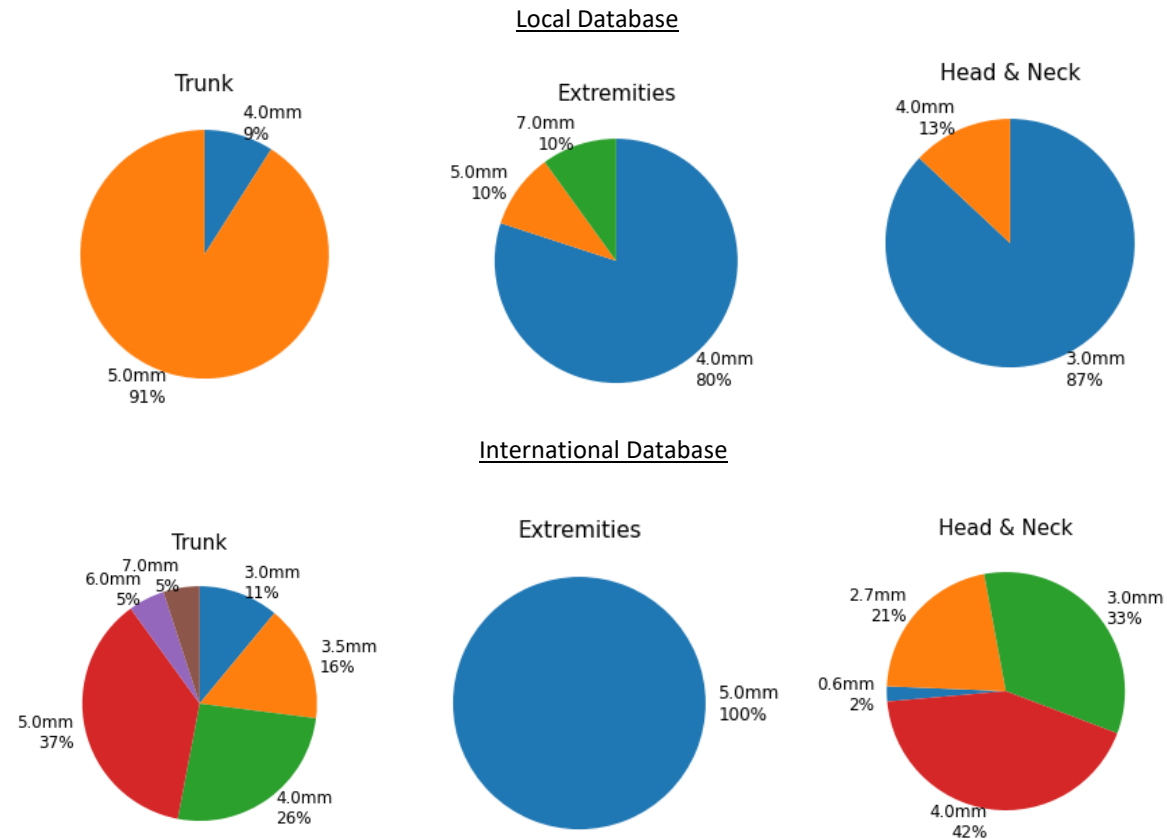


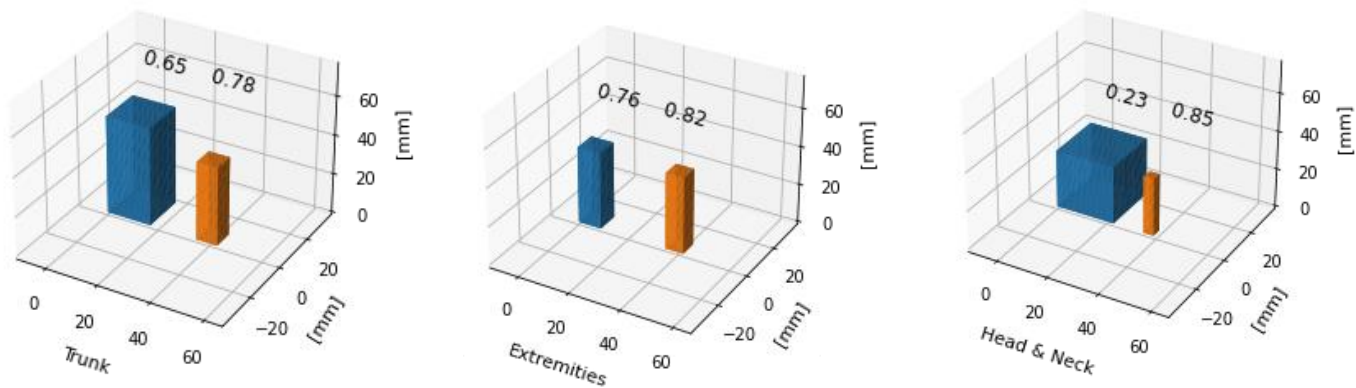
Figure A.7 Pie charts representing the slice thickness for each body part in local and international database.

Table A.2 indicates the results of the ANOVA test on the two databases. The column ‘Size’ indicates the dimension of the 2 or 3 groups compared with each other. The column ‘Available data’ indicates how many values we had for each parameter (max. 17 for local, max. 99 for international database). The ‘lost values’ column indicates the amount of data that did not get included in the two groups (occurrence of the median value). With ‘Diff %’ the percentage difference between the two groups is indicated.

Key	Local Dataset						International Dataset					
	Size	p-val	Aval data	Lost vals	Diff %	Median	Size	p-val	Aval data	Lost vals	Diff %	Median
Patient Age	[7, 7]	0.623	17	3	11%	6	[56, 38]	0.860	77	5	1%	5
Time between scans	[10, 7]	0.017	17	0	41%	50	[64, 35]	5.0e-06	99	0	30%	50
Number of Slices	[8, 7]	0.171	17	2	29%	60	[38, 36]	0.870	99	25	1%	30
Slice Thickness	[3, 5, 9]	0.317	17	0	NaN	[3, 4, 5]	[28, 45]	0.409	99	26	7%	4.0
Spacing betw. Slices	[7, 10]	0.058	17	0	34%	5.0	[46, 53]	0.543	99	0	4%	5.0
Gap	[13, 4]	0.474	17	0	16%	0.0	[20, 79]	0.259	99	0	9 %	0.0
Rows	[7, 10]	0.927	17	0	2%	256	[46, 53]	0.234	99	0	8%	176
Pixel Spacing	[6, 4]	0.118	17	7	44%	1.484	[49, 48]	0.419	99	2	6%	1.471

Anisotropy	[5, 5]	0.037	17	7	53%	0.703	[49, 45]	0.364	99	5	6%	0.677
Volume Pixel	[6, 4]	0.118	17	7	45%	11.017	[49, 44]	0.371	99	6	6%	9.686
Area XY	[8, 8]	0.066	17	1	34%	62500	[36, 48]	0.316	99	15	7%	52900
Volume Slice	[8, 8]	0.131	17	1	30%	264500	[49, 48]	0.919	99	2	1%	242688
Volume Tumor [vox]	[8, 8]	0.526	17	1	12.%	12453	[49, 49]	0.250	99	1	8%	1446
Volume Tumor [mm]	[8, 8]	0.868	17	1	4%	116040	[49, 49]	0.906	99	1	1%	10122
Echo Time	[8, 8]	0.388	17	1	18%	77.80	[52, 47]	0.691	94	0	3%	76.63
Repetition Time	[8, 8]	0.091	17	1	327%	2833.26	[52, 47]	0.332	94	0	7%	4685.86
Diffusion b-value	[1, 3]	0.076	4	13	NaN	[800, 1000]	[1, 1, 7]	0.689	9	90	NaN	[50, 800, 1000]
Magnetic Field	[14, 3]	0.412	17	0	NaN	[1.5, 3.0]	[91, 8]	0.667	99	0	NaN	[1.5, 3.0]

Local Database



International Database

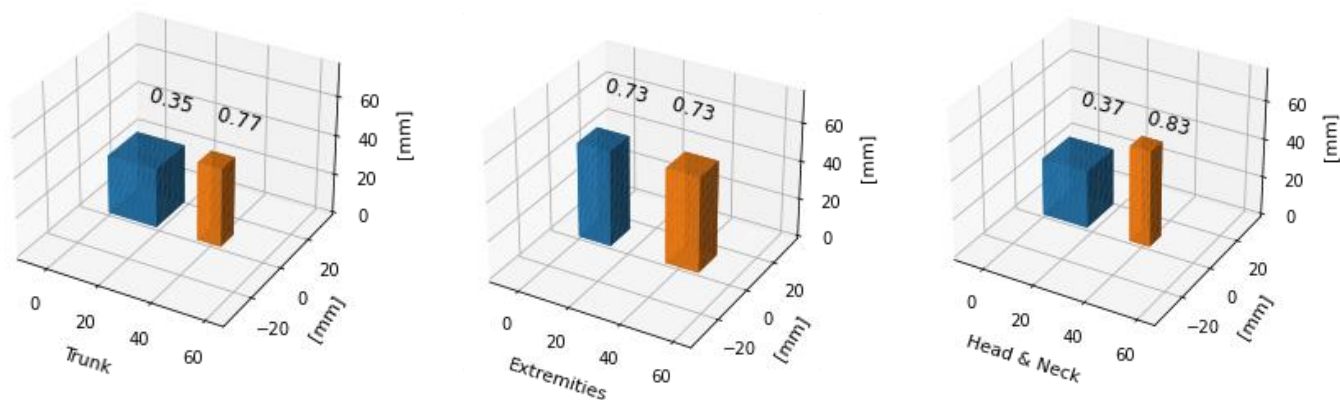


Figure A.8 Tridimensional in-scale representation of the voxel with the maximum (left) and minimum (right) anisotropy ratio in both datasets.

Bibliography

1. Sultan I, Qaddoumi I, Yaser S, Rodriguez-Galindo C, Ferrari A. Comparing Adult and Pediatric Rhabdomyosarcoma in the Surveillance, Epidemiology and End Results Program, 1973 to 2005: An Analysis of 2,600 Patients. *JCO*. 2009;27(20):3391-3397. doi:10.1200/JCO.2008.19.7483
2. van Ewijk R, Vaarwerk B, Breunis WB, et al. The Value of Early Tumor Size Response to Chemotherapy in Pediatric Rhabdomyosarcoma. *Cancers*. 2021;13(3):510. doi:10.3390/cancers13030510
3. Sun X, Guo W, Shen JK, Mankin HJ, Hornicek FJ, Duan Z. Rhabdomyosarcoma: Advances in Molecular and Cellular Biology. *Sarcoma*. 2015;2015:232010. doi:10.1155/2015/232010
4. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*. 2022;72(1):7-33. doi:10.3322/caac.21708
5. Gaal JCV, Teerenstra S, Suurmeijer AJH, Flucke UE, Sleijfer S. The Impact of Age on Outcome of Embryonal and Alveolar Rhabdomyosarcoma Patients. A Multicenter Study. *ANTICANCER RESEARCH*. Published online 2012:13.
6. Quaglia MPL, Heller G, Ghavimi F, et al. The effect of age at diagnosis on outcome in rhabdomyosarcoma. *Cancer*. 1994;73(1):109-117. doi:10.1002/1097-0142(19940101)73:1<109::AID-CNCR2820730120>3.0.CO;2-S
7. van Ewijk R, Schoot RA, Sparber-Sauer M, et al. European guideline for imaging in paediatric and adolescent rhabdomyosarcoma — joint statement by the European Paediatric Soft Tissue Sarcoma Study Group, the Cooperative Weichteilsarkom Studiengruppe and the Oncology Task Force of the European Society of Paediatric Radiology. *Pediatr Radiol*. 2021;51(10):1940-1951. doi:10.1007/s00247-021-05081-0
8. Tong Q, He H, Gong T, et al. Multicenter dataset of multi-shell diffusion MRI in healthy traveling adults with identical settings. *Scientific Data*. 2020;7(1):1-7.
9. Pinto MS, Paoletta R, Billiet T, et al. Harmonization of brain diffusion MRI: Concepts and methods. *Frontiers in Neuroscience*. Published online 2020:396.
10. Stadler A, Schima W, Ba-Ssalamah A, Kettenbach J, Eisenhuber E. Artifacts in body MR imaging: their appearance and how to eliminate them. *European radiology*. 2007;17(5):1242-1255.
11. Afaq A, Andreou A, Koh D. Diffusion-weighted magnetic resonance imaging for tumour response assessment: why, when and how? *Cancer Imaging*. 2010;10(1A):S179.