Cancer imaging and image analysis methods in whole-body MRI and PET/MRI

THERESE SJÖHOLM
Dissertation presented at Uppsala University to be publicly examined in Föreläsningssalen, Röntgen, Akademiska sjukhuset, ingång 70, Friday, 17 March 2023 at 13:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Associate Professor Einar Heiberg Brandt (Cardiac MR group, Lund University).

Abstract

Diagnostic medical imaging techniques such as magnetic resonance imaging (MRI) and positron emission tomography (PET) can provide structural and functional assessments of the whole body. This has great value for potentially systemic diseases such as cancer. To take advantage of the enormous amount of data provided by current imaging systems, improvements in whole-body imaging protocols and advancements in image analysis methods are however needed. This thesis aims to develop advanced imaging and image analysis methods for the purpose of tumour characterisation in MRI and combined PET/MRI whole-body image datasets. Early prediction of progression free survival (PFS) and overall survival (OS) in patients with relapsed/refractory (r/r) large B-cell lymphoma (LBCL) undergoing chimeric antigen receptor (CAR) T-cell therapy was assessed using whole-body PET/MRI pre- and post-therapy. Reference standard manual segmentations of tumours and non-malignant lymphoid tissue were used, and an extended set of semi-quantitative and quantitative PET/MRI metrics was extracted. Predictive PET/MRI metrics included the metabolic tumour volume (MTV), tumour apparent diffusion coefficient (ADC) and 18F-fluorodeoxyglucose (FDG) uptake in non-malignant bone marrow. To enable automated image analysis, deformable image registration was used to create multiparametric normal atlases of healthy volunteers examined with whole-body FDG PET, diffusion weighted imaging (DWI) MRI and water-fat MRI. To improve the geometric accuracy of DWI in the normal atlas, the reverse polarity gradient (RPG) distortion correction method was evaluated. RPG increased the geometrical alignment between DWI and structural images acquired in the same scan session, with little effect on healthy tissue ADC. It was further shown that healthy tissue assessments in atlas space was possible, with the normal atlas employed to study voxel-wise correlations between ADC and age across the whole body, confirming results from a manual segmentation approach. As proof of concept, a probabilistic atlas based approach was successfully used for segmentation of suspected malignant disease in FDG PET data and detection of liver fat infiltration in fat fraction (FF) MRI data. Lastly, using a cohort of r/r LBCL patients, statistical deviations between patient and normal atlas DWI data included as input in a deep learning based model, improved its performance for automated tumour segmentation.

Keywords: magnetic resonance imaging, diffusion weighted imaging, positron emission tomography, whole-body imaging, normal atlas, CAR T-cell therapy, lymphoma, medical image analysis

Therese Sjöholm, Department of Surgical Sciences, Radiology, Akademiska sjukhuset, Uppsala University, SE-75185 Uppsala, Sweden.

© Therese Sjöholm 2023

ISSN 1651-6206
URN urn:nbn:se:uu:diva-495136 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-495136)
Till kvinnorna jag saknar:
mormor Gun, farmor Ethel och svärmor Berit
This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


IV. Sjöholm, T., Tarai, S., Malmberg, F., Strand, R., Korenyushkin, A., Enblad, G., Ahlström, H., Kullberg J. A whole-body diffusion MRI normal atlas: development, evaluation and initial use. *Manuscript*

*Equal contributions as first authors, **equal contributions as senior authors

Reprints were made with permission from the respective publishers.

The author has made substantial contributions to all papers. In paper I, equal contribution was made with Alexander Korenyushkin. The author of this thesis was responsible for quantitative data analysis and writing of the manuscript, excluding the clinical parts on lymphoid tissue assessments in the discussion. The first authors had joint responsibility of interpretation of results. In paper II, equal contribution was made with Simon Ekström. The author was responsible for writing of the manuscript and all aspects of the data analysis excluding image registration. The first authors had joint responsibility of interpretation of results. In papers III and IV the author was the main contributor and responsible for recruitment and data collection of healthy volunteers, data analyses, interpretation of results and writing of manuscripts.
Contents

Introduction ................................................................................................... 11
1.1 Aim ..................................................................................................... 13
1.2 Disposition ......................................................................................... 13

Lymphoma .................................................................................................... 15
2.1 Epidemiology ..................................................................................... 15
2.2 Treatment ........................................................................................... 15

Medical imaging ........................................................................................... 18
3.1 Magnetic resonance imaging .............................................................. 18
   3.1.1 Water-fat MRI ............................................................................ 21
   3.1.2 Diffusion weighted imaging ....................................................... 22
3.2 Positron emission tomography ........................................................... 26
   3.2.1 18F-Fluorodeoxyglucose .............................................................. 26
3.3 Hybrid imaging .................................................................................. 27

Cancer imaging and therapy response assessment ........................................ 29
4.1 Tumour imaging ................................................................................. 29
   4.1.1 Imaging in lymphoma ................................................................. 33
4.2 System level imaging .......................................................................... 34
4.3 Therapy response assessment ............................................................. 34
   4.3.1 Clinical endpoints ....................................................................... 35
   4.3.2 Imaging biomarkers .................................................................... 36

Medical image processing ............................................................................. 37
5.1 Image registration ............................................................................... 37
   5.1.1 Imiomics ..................................................................................... 39
5.2 Tumour segmentation ......................................................................... 42
   5.2.1 Whole body applications ............................................................ 44

Contributions ................................................................................................ 46
6.1 Paper I ................................................................................................ 47
6.2 Paper II ............................................................................................... 50
6.3 Paper III .............................................................................................. 52
6.4 Paper IV .............................................................................................. 55

Discussion ..................................................................................................... 58
7.1 Imaging contrasts ................................................................................. 58
Abbreviations

2D  Two-dimensional
3D  Three-dimensional
ADC  Apparent diffusion coefficient
AP  Anterior-posterior
BMI  Body mass index
CAR  Chimeric antigen receptor
CNN  Convolutional neural network
CMR  Complete metabolic response
CT  Computed tomography
DLBCL  Diffuse large B-cell lymphoma
DWI  Diffusion weighted imaging
DWIBS  Diffusion weighted whole-body imaging with background body signal suppression
EPI  Echo planar imaging
FDG  $^{18}$F-fluorodeoxyglucose
FDR  False discovery rate
FF  Fat fraction
FID  Free induction decay
FOV  Field of view
GE  Gradient echo
HR  Hazard ratio
iRECIST  immune RECIST
LBCL  Large B-cell lymphoma
MI  Mutual information
MRI  Magnetic resonance imaging
MTV  Metabolic tumour volume
NMR  No metabolic response
OS  Overall survival
PA  Posterior-anterior
PE  Phase encoding
PERCIST  PET response criteria in solid tumors
PET  Positron emission tomography
PFS  Progression free survival
PMD  Progressive metabolic disease
PMR  Partial metabolic response
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIBA</td>
<td>Quantitative imaging biomarkers alliance</td>
</tr>
<tr>
<td>RC</td>
<td>Repeatability coefficient</td>
</tr>
<tr>
<td>RECIST</td>
<td>Response evaluation criteria in solid tumors</td>
</tr>
<tr>
<td>RF</td>
<td>Radiofrequency</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>RPG</td>
<td>Reverse polarity gradient</td>
</tr>
<tr>
<td>r/r</td>
<td>Relapsed/refractory</td>
</tr>
<tr>
<td>SE</td>
<td>Spin echo</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to noise ratio</td>
</tr>
<tr>
<td>STIR</td>
<td>Short TI inversion recovery</td>
</tr>
<tr>
<td>SUV</td>
<td>Standardised uptake value</td>
</tr>
<tr>
<td>T₁</td>
<td>Longitudinal relaxation time</td>
</tr>
<tr>
<td>T₂</td>
<td>Transverse relaxation time</td>
</tr>
<tr>
<td>tDV</td>
<td>Total diffusion volume</td>
</tr>
<tr>
<td>TE</td>
<td>Echo time</td>
</tr>
<tr>
<td>TI</td>
<td>Inversion time</td>
</tr>
<tr>
<td>TNM</td>
<td>Tumor, node, metastasis</td>
</tr>
<tr>
<td>TR</td>
<td>Repetition time</td>
</tr>
<tr>
<td>TLG</td>
<td>Total lesion glycolysis</td>
</tr>
<tr>
<td>TOF</td>
<td>Time of flight</td>
</tr>
<tr>
<td>V</td>
<td>Structural volume</td>
</tr>
<tr>
<td>VOI</td>
<td>Volume of interest</td>
</tr>
<tr>
<td>WF</td>
<td>Water fraction</td>
</tr>
</tbody>
</table>
Introduction

The incidence of cancer has increased continuously during the past decades, partly due to an aging population, life style factors, the introduction of screening programs and improved methods for detection. Although cancer treatments in many cases are successful, and the proportion of cured cancer patients has increased dramatically, cancer is still a serious public health threat and the most common cause of death below the age of 80 years in Sweden [1] and the second leading cause of death worldwide [2]. According to current understanding, cancer is a highly complex disease. The cancer cell population is viewed as heterogeneous, with genetic and epigenetic variations contributing to clonal evolution and tumour heterogeneity, as well as differences in the tissue microenvironment (e.g. blood vessels, immune cells and stromal cells) and inflammatory processes [3,4].

Conventional cancer therapies (surgery, radiotherapy, chemotherapy) were developed with little knowledge of cancer biology. Translational cancer research based on cancer biology was initiated in the 1990s, with the first targeted drugs against a molecular target available in the early 2000s [5]. While chemo- and radiotherapy act on rapidly dividing cells, with no discrimination between healthy and cancerous cells, targeted therapies block tumour growth by interfering with specific molecules involved in cancer progression while sparing off-target cells. Another type of increasingly used targeted therapy is immunotherapy. An immunotherapy drug does not act directly on tumour cells, but instead activates the immune system to recognise and kill tumour cells. Immunotherapy has improved treatment outcomes in many cancers, mainly due to immune checkpoint inhibitors and T-cell transfer therapy [6].

Personalised medicine in the context of cancer therapy can be described as the usage of patient-specific factors, e.g. tumour heterogeneity and microenvironment, lifestyle and co-morbidities, to guide treatment decisions. Due to intra- and inter-tumour heterogeneity, cancer can develop very differently for different patients. Frequent metastatic mutations mean that different cell clones are created, giving a variation in therapy response at the tumour and patient level. As the number of available therapies is constantly increasing, biomarkers that can predict therapy outcome and aid in determining the optimal use of available therapies are needed. A biomarker is defined as ‘a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a
therapeutic intervention’ [7] and can stem from e.g. genes, proteins or radioactively labelled molecules, but also from measurements in diagnostic medical images [8].

Diagnostic medical imaging includes a number of imaging modalities, e.g. computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET). CT, MRI and PET all have the capability to provide three-dimensional (3D) images of the whole body and can be used for tumour staging and longitudinal follow-up. Whole-body imaging could be of great value in oncology since almost all cancers are potentially systemic. Structural whole-body images of tumour and healthy tissue morphology can be obtained by the means of CT and MRI, while physiological whole-body images of typical cancer hallmarks (e.g. cell proliferation, hypoxia, metabolism, apoptosis and angiogenesis) can be made available via contrast-enhanced CT, MRI and PET. In this sense, medical imaging gives the opportunity to study both intra- and inter-tumour structure and function across the whole body.

Medical image interpretation in clinical practice typically depends on visual assessment with an observer using his/her accumulated knowledge to recognise what is normal and abnormal. The extensive amount of data created by modern imaging systems, in particular from the comprehensive protocols of hybrid PET/CT and PET/MRI systems, implies that this observer-driven pattern recognition approach becomes very time-consuming. Automated methods for advanced image analysis are needed to take advantage of the enormous amount of data provided by current imaging systems and to support integrated analysis of available metrics. In cancer, improving the methods for imaging, detection and characterisation of tumours means improved possibilities to evaluate personalised medicine approaches, choosing the best treatment for an individual patient and being able to predict therapy response.

Statistical atlases have been composed for a number of anatomical sites using medical image registration. The predominant field of research has been in brain imaging studies [9], with single- or multi-modal representations of the healthy and diseased brain being constructed and group-specific patterns of brain structure and function being identified. This approach has been less studied on a whole-body scale, largely due to the complexity involved when registering whole-body medical images. Systemic diseases such as cancer could however benefit from a whole-body statistical atlas approach, as it has the potential to provide automated analyses of tissues and tumours across the whole body without confining the analyses to a single tissue and/or a limited number of tumours as traditionally performed.
1.1 Aim

The overall aim of this thesis is to develop imaging and image analysis methods for the purpose of detection and characterisation of tumours in MRI and PET/MRI whole-body image data. Three main specific objectives are identified:

- Perform multi-parametric characterisations of tumours for the purpose of improved diagnostics and therapy evaluations in whole-body datasets
- Create multi-parametric whole-body MRI and PET/MRI atlases of healthy subjects
- Perform automated voxel-wise whole-body image analyses using the multi-parametric atlases and cancer imaging data

1.2 Disposition

This is a comprehensive summary thesis based on four papers as summarised in chapter 6.

Paper I targets semi-quantitative and quantitative PET/MRI metrics of tumour and non-malignant lymphoid tissue for early prediction of progression free survival (PFS) and overall survival (OS) in patients with relapsed/refractory (r/r) large B-cell lymphoma (LBCL) undergoing chimeric antigen receptor (CAR) T-cell therapy.

In Paper II, a multi-parametric whole-body atlas of functional FDG PET and structural water and fat separated (water-fat) MRI data of adult healthy volunteers is developed for the purpose of voxel-based analysis. As proof of concept, the atlas is used for automated voxel-wise anomaly detection.

In Paper III, the performance of the reverse polarity gradient (RPG) method for correction of susceptibility-induced geometric distortion of whole-body diffusion weighted imaging (DWI) at 1.5T and 3T is assessed for healthy volunteers.

In Paper IV, whole-body diffusion atlases of adult healthy volunteers scanned at 1.5T and 3T are developed. The atlases are evaluated by establishing whole-body apparent diffusion coefficient (ADC) values of healthy tissues and by including healthy tissue deviations in an automated tumour segmentation task.

The framing text provides the context of these four papers. Chapter 2 gives a brief introduction to lymphoma and its treatments, with focus on novel CAR T-cell therapy. Chapter 3 describes the basic theory of the medical imaging techniques used in this thesis, including water-fat MRI, DWI MRI and FDG PET, as well as PET/MRI hybrid imaging systems. Chapter 4 gives an overview of cancer imaging and therapy response assessment techniques using...
these imaging modalities, while Chapter 5 introduces medical image processing, with focus on image registration and tumour segmentation. Chapter 6 outlines the contributions of this thesis to cancer imaging and image analysis research, while a discussion is given in Chapter 7. Chapter 8 provides a short summary in Swedish.
Lymphoma

The human immune system is highly complex and comprises two main parts: the innate and the adaptive immune system. The innate immune system provides an antigen-independent rapid and short-lived response to pathogens entering the body. The adaptive immune system is antigen-specific and utilizes lymphocytes in the form of B- and T-cells that proliferate after antigen recognition, destroying antigenic structures. It further provides immunologic memory, with an effective response initiated for repeated exposures to the same pathogen.

Lymphoma is a type of hematologic malignancy that starts in the lymphocytes of the immune system. This chapter provides an overview of its epidemiology and treatment strategies.

2.1 Epidemiology

Lymphoma accounts for approximately 3.2% of all new cancer cases worldwide and has a higher incidence in countries with high/very high human development index [10]. It is a heterogenic cancer with a large number of distinct subtypes [11], with diverse treatment pathways and outcomes. Aggressive lymphomas have a fast clinical progression and in general require immediate treatment. Indolent subtypes tend to have a slow clinical progression, with the possibility to use active surveillance instead of immediate treatment. The majority of lymphomas (95%) stem from B-cells, with the most common types being diffuse large B-cell lymphoma (DLBCL) (30-40%) and follicular lymphoma (20%) [12].

2.2 Treatment

Improvements in treatment regimens have meant a continuously increasing trend in the 5- and 10-year survivals of lymphoma patients (Figure 2.1). Largely, this is due to the addition of the anti-CD20 monoclonal antibody rituximab to different chemotherapies. In DLBCL, combination treatment of chemotherapy and rituximab now mean complete responses are achieved in 70-80% of patients. 20-30% of patients are however refractory (i.e. the disease
no longer responds to therapy) or relapse within five years (i.e. the disease return after treatment) [13]. For younger and fit patients with r/r disease (<70 years, higher performance status, lower comorbidity load), curative second-line treatment is possible with high dose chemotherapy followed by autologous stem-cell transplantation. For non-responders and patients not eligible for stem cell transplantation, alternative treatments are needed [14].

Figure 2.1. 5- and 10-year relative survival for lymphoma (excluding Hodgkin lymphoma) in Sweden between 1980 and 2016. Adapted from the Swedish Cancer Society [15].

**CAR T-cell therapy**

One of the hallmarks of cancer cells is avoidance of immune destruction [16]. In cancer immunotherapy, this hallmark is targeted by activating the inherent mechanisms of the immune system to recognise and destroy tumour cells. CAR T-cell therapy is a type of adoptive T-cell immunotherapy in which genetically modified T-cells are manufactured and made to recognise specific surface antigens on tumour cells. The therapy is a multi-step process as illustrated in Figure 2.2. T-cells are first collected from the patient, followed by incorporation of CAR into the T-cell genome and *in vitro* expansion. The patient receives lymphodepleting chemotherapy to make room for the CAR T-cells, which are then infused into the patient. Once infused, the CAR T-cells
recognise tumour target antigens and proliferate. Several CAR T-cell therapies have gained approval in the United States and Europe for usage in B-cell lymphoma, all targeting the CD19 cell surface antigen [17].

Although durable remissions with complete response rates of 40-59% have been reported in clinical trials [18], many patients eventually relapse after CAR T-cell therapy. Limitations include antigen escape, poor T-cell persistence and life-threatening side-effects such as cytokine release syndrome and neurotoxicity [19]. In addition to improvements of the CAR T-cell therapy, biomarkers that can predict durable responses and early recognise non-responders are needed [20].

Figure 2.2. An illustration of the CAR T-cell therapy multi-step process: Lymphocytes are obtained from the patient and T-cells extracted. These T-cells are genetically modified with CAR and thereby transformed into CAR T-cells. CAR T-cells are expanded \textit{in vitro} and then infused into the patient. VL, variable region of light chain; VH, variable region of heavy chain. (From: “Current progress in CAR-T cell therapy for tumor treatment”, by Chen et al, Oncol Lett 2022 24:358 [21]. Licence: CC BY-NC-ND 4.0)
Medical imaging refers to a vast number of imaging techniques used to visualise body tissue and organs, often non-invasively. It is an important component in the cancer care pathway, with applications in disease diagnosis, treatment follow-up and monitoring over time.

In this thesis, MRI and PET whole-body medical imaging has been used. These techniques are briefly described in this chapter, together with the imaging contrasts employed: water-fat MRI, DWI and 18F-Fluorodeoxyglucose (FDG) PET. The chapter ends with a short discussion on hybrid imaging systems.

3.1 Magnetic resonance imaging

In MRI, magnetic resonance is utilised to image the spatial localisation of nuclei with non-zero spin quantum numbers, i.e. with an associated magnetic moment, within the body. Due to its high natural abundance in the body and its large magnetic moment, the 1H nucleus (i.e. proton) with half-integer spin, is the most common nucleus used in MRI. Imaging is performed with three different types of magnetic fields; a static magnetic field ($B_0$), a radiofrequency (RF) field and time-varying magnetic field gradients.

Nuclei with non-zero spin quantum numbers effectively behave like microscopic magnetic dipoles and as a result, two effects will occur as these nuclei are exposed to $B_0$. Firstly, the spins will start to precess around the direction of $B_0$. The precession frequency is given by the resonance frequency (also called Larmor frequency) according to Equation 3.1.

$$\omega_0 = \gamma B_0$$  \hspace{1cm} (3.1)

In which $\omega_0$ is the angular frequency of the spins. For $^1$H, the resonance frequency is 42.58 MHz/T. Different types of nuclei have different gyromagnetic ratios ($\gamma$), meaning the precession frequency will differ between isotopes. Secondly, the direction of the nuclear magnetic moments, coinciding with their spin direction, tend to align with $B_0$. As a result, a net magnetisation aligned parallel with $B_0$ will be created. The magnitude of the net magnetisation differs between different tissues depending on their proton density.
Nuclear magnetic resonance is the phenomenon occurring when the normal precession of a nucleus, in a static magnetic field, is increasingly perturbed as a result of an external, relatively weak, magnetic field applied at its precession frequency. In MRI, this perturbation is obtained with an RF pulse (i.e. an RF field of short duration, also called excitation). As an RF pulse is applied at the resonance frequency, the protons absorb energy from the pulse, effectively tipping the net magnetisation from its alignment with $B_0$ into the transverse x-y plane of the scanner. The x- and y-axes of the magnet are defined in the left-right and top-bottom directions of the bore (transverse plane), whereas the z-axis is defined in the same direction as $B_0$. To tip the net magnetisation exactly into the transverse plane, a 90° RF pulse is required. The flip angle ($\alpha$) is determined by $\gamma$, the strength of the applied RF pulse ($B_1$) and the duration of the RF pulse ($t_p$) according to Equation 3.2.

$$\alpha = \gamma B_1 t_p \quad (3.2)$$

A second effect of the RF pulse is to bring spins into phase coherence. In this way, the applied RF excitation pulse creates a transverse magnetisation vector. After an RF pulse has been applied, the transverse magnetisation vector rotates about the axis of $B_0$ at the resonance frequency. This effectively produces a rotating magnetic field, which can be sampled by receiver coils in which electric currents, representing the MR signal, are induced.

The gradient system is used to spatially localise the MR signal. This is performed by using gradient coils creating magnetic field gradients in three orthogonal directions; $G_x$, $G_y$ and $G_z$. Gradient amplifiers supply the gradient coils with electrical current, ideally giving rise to trapezoidal gradient pulses. The gradient amplitude is the strength of the gradient, which has a maximum value in the range of 10-50 mTm$^{-1}$. Higher gradient amplitudes enables thinner slice thicknesses and smaller field of views (FOVs) to be set. A gradient cannot be increased to its full strength instantaneously, but needs a rise time. The rise time is usually given in $\mu$s and typical values are in the range of 200-1000 $\mu$s. By dividing the gradient amplitude with the rise time, the slew rate of the gradient system is obtained. High slew rates are necessary for fast imaging sequences as it sets the minimum echo time (TE) and repetition time (TR) possible to achieve (see Figure 3.1 for definitions of TE and TR). Spatial encoding of the MR signal is performed with the gradients by producing a variation in $B_0$ in the x, y and z directions of the scanner. To spatially encode the MR signal, slice select, phase encode and frequency encode gradients are used. These gradients can be applied in any direction by combining the orthogonal x, y and z gradients. The slice select gradient is switched on as an RF pulse is applied, to only excite tissue within a particular slice of the body. To localise the signal in the two remaining orthogonal directions within the slice, the phase and frequency encode gradients are utilised. Phase encoding
(PE) is performed prior to data acquisition to encode the phase of the spins. A frequency encode gradient is applied as the MR signal is recorded to encode the signal in terms of frequency. The acquired signal is hence encoded in phase and frequency. Inverse Fourier transformation is used to reconstruct the MR signal in time domain.

**Signal Generation**

As an RF excitation pulse is applied, a component of the longitudinal magnetisation aligned with $B_0$ is tipped into the transverse plane. When the pulse is switched off, an exponentially decaying signal, called the free induction decay (FID), is created as illustrated in Figure 3.1. The FID is an example of spin-spin relaxation, resulting from gradual loss of spin phase-coherence after the RF pulse is applied. The rapid decay of the FID and the fact that its signal is at its maximum immediately after the excitation pulse make it difficult to measure and more seldom used for signal acquisition. Instead, acquisition is most commonly performed with two different types of echoes; spin echo (SE) or gradient echo (GE). In a SE sequence, each slice of tissue is excited with a 90° RF pulse, followed by a 180° refocusing pulse as illustrated in Figure 3.1. In a GE sequence, an excitation pulse with a typically lower flip angle is used, followed by refocusing with magnetic gradients prior to readout.

![Figure 3.1](image)

**Figure 3.1.** Left: The FID curve. The signal decays with time constant $T_2^*$, which depends on $B_0$ inhomogeneity and spin-spin relaxation. Right: Signal generation for a basic SE sequence. A slice selective 90° excitation pulse is followed by a 180° refocusing pulse. TE is the time between the excitation pulse and the refocused spin echo. TR is the time between two successive 90° excitation pulses.

The signal obtained in MRI comes from the rotating transverse magnetisation vector defined by the two properties magnitude and direction. Independent of sequence used, conventional MRI uses the former and displays the modulus of the signal. Brighter pixels in the images hence correspond to stronger signals. The MRI signal is not quantitative, but more advanced sequences can be used for this purpose, e.g. water-fat and diffusion MRI as described below.
3.1.1 Water-fat MRI

Structural MR images used in this thesis were acquired using chemical shift-based water and fat separation methods. The MR signal from $^1$H in the human body mainly stems from water, but a large portion of the signal also derives from fat. Due to magnetic properties of surrounding atoms, the protons of water and of the triglyceride chains of fat experience slightly different magnetic field strengths and as such have slightly different resonance frequencies. This effect is called chemical shift and can be used to create chemical shift-based water and fat separated images (in this work called water-fat MRI).

Water-fat MRI was first described by Dixon et al [22] and methods for producing these types of images are hence often referred to as Dixon methods. These methods are able to separate the water and fat signals by taking advantage of the phase shifts created as a result of the resonance frequency differences between water and fat. By acquiring images at different TEs, phase information can be encoded. In its simplest form, water-fat MRI is acquired using a two-point method; images with two different TEs are acquired for which the water and fat signals are in-phase ($S_{in}$) and out-of-phase ($S_{out}$) in the transverse plane. After RF excitation, the water and fat signals are opposed at 2.3 ms and aligned at 4.6 ms at 1.5T. At 3T, this timing is halved. Water (W) and fat (F) images are separated after acquisition (Equations 3.3 and 3.4). From reconstructed water and fat images, it is possible to quantify the fat fraction (FF) and water fraction (WF) voxel contents (Equations 3.5 and 3.6) [23].

$$W = (S_{in} + S_{out})/2 \quad (3.3)$$

$$F = (S_{in} - S_{out})/2 \quad (3.4)$$

$$FF = F/(W + F) \quad (3.5)$$

$$WF = W/(W + F) \quad (3.6)$$

The original two-point method is sensitive to $B_0$ inhomogeneity, giving water-fat swaps in reconstructed images. This prompted the development of more advanced water-fat imaging methods, including three- and four-point approaches, and more advanced reconstruction algorithms [24]. Water-fat imaging methods used in this work are all two-point Dixon GE sequences, with a typical example shown in Figure 3.2.
3.1.2 Diffusion weighted imaging

Using DWI, the random translational motion of molecules (Brownian motion) can be imaged. Typically, the MRI signal from fat is suppressed to only target the motion of water molecules. Free diffusion relates to uninhibited motion of water molecules. In tissue, motion is however restricted as the water molecules bounce, cross and interact with tissue components such as cell membranes, fibres and macromolecules. DWI can hence provide information about structural features and geometric organisation of tissue on a microscopic scale despite a spatial resolution of a substantially coarser scale. Technically, the MRI signal is made sensitive to diffusion by adding diffusion sensitising gradients to a standard MRI sequence. The same gradient coils as used for spatial encoding generate the diffusion gradients. This is illustrated in Figure 3.3 for a SE sequence.

Figure 3.3. Example of a SE diffusion sensitised sequence. Symmetric diffusion gradients are added on both sides of the 180° refocussing RF pulse. The gradient duration (\( \delta \)), amplitude (G) and separation (\( \Delta \)) are set to make the sequence sensitive to diffusion of a particular size. The diffusion sensitising gradients can be applied to the x-, y-, or z-axes (G_{xyz}), or to a combination of them (the diffusion sensitising direction).
The sensitivity of the diffusion sequence to water motion can be tuned by changing the sequence diffusion-weighting factor (b-value), where a higher b-value corresponds to imaging with emphasis on slower diffusion (i.e. imaging of slow moving water molecules and smaller diffusion distances). The b-value depends on the time between diffusion gradients ($\Delta$), gradient amplitude (G) and gradient duration ($\delta$) according to Equation 3.7.

$$b = \gamma^2 G^2 \delta^2 \left( \Delta - \frac{\delta}{3} \right) \quad (3.7)$$

In practice, more than one b-value is acquired per imaging volume to create images sensitive to different levels of diffusion. By acquiring more than one b-value, the diffusion can be quantified according to Equation 3.8, in which $S_b$ is the signal measured at a set b-value, $S_0$ is the signal measured without diffusion sensitising gradients and ADC is the apparent diffusion coefficient.

$$S_b = S_0 e^{-b*ADC} \quad (3.8)$$

Clinical DWI is typically conducted with a single shot echo planar imaging (EPI) acquisition, which has the advantage of fast signal recording, meaning macroscopic motion is minimised. Disadvantages mainly include image distortions due to susceptibility effects, a low bandwidth in the PE direction and eddy currents (and to a lesser extent also poor spatial resolution and T2 blurring) [25].

For oncological applications, whole-body DWI with background body signal suppression (DWIBS) has been implemented [26,27]. Using DWIBS, background body signal suppression of vessels, muscle and fat is obtained by using high b-values ($b \approx 800-1000 \text{ s/mm}^2$) and short TI inversion recovery (STIR) fat suppression. TI stands for inversion time, i.e. the time between a 180° inversion RF pulse and a 90° excitation RF pulse, set to null the MRI signal from fat. The acquisition is performed in free breathing to allow for multiple signal averaging. DWIBS can be performed at both 1.5T and 3T. Although an increased signal-to-noise ratio (SNR) is obtained at 3T, the drawback is increased susceptibility artefacts (giving distortion and/or signal dislocation) and difficulty in getting a uniform fat suppression [28]. For ADC quantification in whole-body DWI, it is recommended to use at least two b-values; a b-value of 50-100 s/mm$^2$ to suppress intravascular signal and a b-value of 600-1000 s/mm$^2$ to suppress signal from normal tissues [29]. Example whole-body DW images are shown in Figure 3.4.
Figure 3.4. Example of coronal whole-body DW images (b=50, 400 and 900 s/mm²) and corresponding ADC map for a patient with lymphoma. The DW images are shown in inverted grey scale. Due to normal tissue signal suppression, tumours are more clearly visible on the b=900 s/mm² image compared to lower b-value images. In the ADC map, malignant tumours have a dark appearance due to increased cell density. au, arbitrary units.

**Image distortion and correction strategies in DWI**

In DWI, image distortion is common due to both gradient hardware and sequence-specific limitations, and varies from subject to subject. In general, this distortion means that DW and structural MR images acquired in the same scan session will not be geometrically aligned (see example in Figure 3.5). In multiparametric datasets, this mismatch can lead to difficulties in segmentation and quantification. For integration of DWI into workflows for large-scale analysis of multi-parametric oncology data, geometric alignment between input images is essential.

The main source of distortion due to gradient hardware is eddy currents. These occur due to the strong and rapidly switching gradient pulses used in DWI, with induction of electric currents in conducting surfaces as a result. The eddy currents generate magnetic field gradients that combine with the MR gradient pulses and create image distortion. Distortion will vary depending on gradient sensitisation (b-value). As several b-values are used for ADC calculation, this will result in blurred and inaccurate quantitative DWI. Due to a mismatch between actual and assumed local gradient strength, the voxel size can also vary slightly depending on gradient sensitisation. To decrease eddy currents, modern scanners use actively shielded gradients. Eddy currents can also be limited by selecting appropriate gradient waveforms (e.g., pre-emphasis to oppose eddy currents) or by correcting the k-space data (e.g., by calibrating eddy-current artefacts in k-space) [30]. In addition, image registration
have been used for correcting distortion due to eddy currents in brain applications. Through various image registration techniques, non-corrected DW images are registered to less distorted images (e.g. structural MRI) [30–32].

The largest source of distortion in whole-body DWI stems from the sensitivity of the EPI-based acquisition to $B_0$ inhomogeneity (Figure 3.5). $B_0$ inhomogeneity mainly occurs due to magnetic field susceptibility differences between tissues and at tissue air transitions [25]. This distortion cannot be accurately modelled for or shimmed in advance as it is subject-specific; it depends on magnetic susceptibilities within the subject and the geometry of the subject. In multi-station acquisitions such as those used in whole-body DWI, $B_0$ inhomogeneity can also give central frequency shifts between acquired stations. Several methods for retrospectively correcting diffusion images for distortion have been proposed, all aiming to measure $\Delta B_0$ by acquiring additional sequence/s. Example methods include residual field maps [33], point spread function approaches [34] and usage of an RPG acquisition [35]. The latter was assessed for whole-body imaging in Paper III of this thesis.

In contrast to eddy currents, the susceptibility-induced distortion is typically the same regardless of diffusion-weighting factor, meaning the same distortion pattern is observed for all acquired b-values. Its effect is more pronounced along the PE direction compared to the frequency encoding direction (due to the shorter bandwidth of the former), and at higher field strengths [36].

Figure 3.5. Examples of distortion due to $B_0$ inhomogeneity in the brain (left) and whole body (right). DW images acquired with a b-value of 0 s/mm$^2$ are shown in inverted grey scale together with corresponding structural MR images. The distortion can be seen to occur in the PE direction as indicated by arrows. For the whole-body sagittal image the distortion is mainly visible at station boundaries.
3.2 Positron emission tomography

In PET, the focus is to image e.g. organ function and receptor expression, looking at the distribution of a specific radiotracer in the body at a fixed time point (static imaging) or over time (dynamic imaging). Anatomical information is to some extent also obtained, but with inferior spatial resolution compared to MRI and CT.

PET imaging is based on mutual annihilation of a positron and an electron, giving emission of two 511 keV photons. The annihilation occurs when the positron and electron essentially are at rest, meaning the total emission energy will equal the rest mass energies of the two particles and the created photons will in total have zero momentum. The photons will hence travel in opposite directions from the interaction site. In PET, the patient is surrounded by a ring of detectors, meaning the almost simultaneous arrival of the photons can be recorded. Using techniques such as coincidence timing and time of flight (TOF), the approximate location of each annihilation event can be determined.

After data acquisition, images are reconstructed. Iterative reconstruction is currently used clinically, with corrections (e.g. normalisation, dead-time, scatter, attenuation) incorporated into the reconstruction algorithm. The resulting tomographic images have a voxel count rate that corresponds to the radioactivity concentration. Hence, PET is referred to as a quantitative imaging technique.

Clinically, the standardised uptake value (SUV) is commonly used to specify the amount of radiotracer uptake in tissue (Equation 3.9). ‘Body size’ is the volume of distribution and is commonly represented by the total body weight or lean body mass. A major advantage of the SUV metric compared to a full kinetic modelling approach is that only a static scan is needed, compared to the complex dynamic studies, often requiring arterial blood sampling, needed for kinetic modelling. The drawback is however that a semi-quantitative measure is obtained, as the SUV depends on a number of factors, e.g. reconstruction parameters, scanner calibration and physiological factors [37].

\[
SUV = \frac{\text{activity concentration in tissue}}{\text{injected activity/body size}} \quad (3.9)
\]

3.2.1 $^{18}$F-Fluorodeoxyglucose

In PET, molecules labelled with radioactive nuclides (radiotracers) are used. The administered amount of a radiotracer is very small, as to not exhibit any pharmacological effect. Molecular changes can therefore be assessed without changes to the molecular or biochemical process being studied. Although a number of radiotracers have been developed for usage in oncology [38], FDG is by far the most common. FDG is a glucose analogue transported into cells by facilitated diffusion and then trapped intra-cellularly. Malignant tumours
are characterised by an increased glucose transport, with the uptake into tissue being proportional to the overall glucose metabolism. FDG uptake by malignant tumours is however non-specific as the radiotracer is also taken up by inflammatory tissue. A high uptake of FDG is normally obtained in the brain, kidneys and liver, as well as in the bladder due to its excretion route (Figure 3.6).

There are a number of practical considerations to take into account to achieve good image quality when imaging with FDG. Prior to injection, the subject should fast for a set amount of time (normally >6 h). Low insulin levels are in this way obtained, yielding less uptake into muscle and fat, and a better target to background ratio. The glucose level is hence routinely checked prior to the FDG injection. After intravenous injection of FDG, the subject should preferably lie down and relax until being scanned. This to avoid a high tracer uptake in muscles. A comfortable temperature should be maintained during the rest period in order to avoid undesirable uptake in brown adipose tissue. The rest period between injection and scan time should be standardised. Imaging at the same time after tracer injection is important for accurate comparisons between studies.

Figure 3.6. Example of a whole-body FDG PET image for a patient with lymphoma. An increased FDG uptake is seen for more metabolically active tissues, such as brain, liver and tumours. Uptake is also seen in the kidneys, as FDG is excreted via urine.

### 3.3 Hybrid imaging

Originally, imaging modalities such as PET, CT and MRI were used separately in the diagnosis, staging and monitoring of disease. Hybrid PET/CT imaging introduced in the early 2000s, however, enabled sequential PET and CT scans to be performed in the same scanning session, minimising temporal and
spatial differences between the two sets of images [39]. Inherently co-regis-
tered functional and structural images can in this way be obtained, which has
proven useful in a number of clinical oncology applications [40].

Hybrid PET/MRI systems were more recently introduced and although still
primarily applied in research, clinical applications are becoming more com-
mon [41]. Compared to sequential PET/CT, integrated PET/MRI offers truly
simultaneous PET and MR imaging. As MRI does not involve ionising radia-
tion, PET/MRI gives a reduced radiation exposure compared to PET/CT. In
oncology, the combination of MR multi-parametric imaging features and PET
tracer molecular information is promising, both in terms of providing detailed
information of lesions across the whole body and also for assessment of lesion
heterogeneity [42]. Disadvantages of PET/MRI compared to PET/CT include
e.g. the longer examination time for MRI compared to CT and the lack of
directly measurable attenuation coefficients from MRI.
Cancer imaging and therapy response assessment

MRI and PET/MRI systems can collect a large number of complementary imaging contrasts and are used in all phases of cancer management, from early detection via screening to post-therapy disease monitoring. Different aspects of tumour growth can be mapped, from the macroscopic level of tumour structure to the microscopic level of e.g. tumour diffusion, perfusion and metabolism. In addition to qualitative image analysis, semi-quantitative and quantitative metrics can be extracted to aid in image interpretation. The latter is particularly useful for objective assessment of tumours over time.

The tumour imaging contrasts used in this work are described in this chapter together with commonly used evaluation metrics and a brief overview of whole-body imaging in lymphoma. In addition to tumour assessments, whole-body imaging can provide a system level type of analysis as briefly described in Chapter 4.2. The chapter ends with an overview of concepts related to cancer therapy response assessment.

4.1 Tumour imaging

Water-fat MRI

The structural images used in this work stem from water-fat MRI. Water-fat MRI can provide qualitative assessment of the tumour and its environment from four imaging contrasts (in-phase, out-of-phase, water-only and fat-only images as partly illustrated in Figure 3.2), and quantitative measurements of tumour size and fat infiltration (FF, Equation 3.5). Tumour size is a widely used metric for staging and, as further discussed in Chapter 4.3, therapy response assessment. Currently, unidimensional measures of tumour size, such as the largest diameter, are widely used and well validated [43]. FF has been shown particularly useful as a quantitative marker of malignant infiltration in bone in multiple myeloma, for which the bone marrow composition, i.e. the proportion of haematopoietic (red) and fatty (yellow) marrow, is altered [44,45].
**Diffusion weighted imaging**

In clinical care, DWI has most extensively been used in brain ischemia. The technique is however increasingly used in oncology. As compared to healthy tissues, most malignant tumours exhibit an impeded diffusivity due to increased cellularity, with a relatively high signal intensity seen on high b-value DW images. Necrotic and apoptotic processes can however lead to a counteracting increase in diffusivity due to loss of cell membrane integrity and a decrease in cellularity. In a heterogeneous tumour, a more cystic or necrotic fraction of the tumour will hence show greater signal attenuation on high b-value images, as water diffusion is less restricted [46]. DWI can hence be useful for both tumour detection and characterisation and has shown promising results in many applications [47].

![Diagram](image)

**Figure 4.1.** A schematic illustration of the change in cellularity (left) and corresponding increase in molecular water mobility, measured by the ADC (right), as a tumour responds to treatment (from top to bottom). For a tumour responding to therapy, an increase in extracellular space/membrane permeability allows for greater water mobility and increased ADC. (From: “Diffusion magnetic resonance imaging: a biomarker for treatment response in oncology”, by Hamstra et al, J Clin Oncol. 2007 25:26 [48]. Reprinted with permission from Journal of Clinical Oncology.)

When tumours are treated, therapy-induced cell death by apoptosis, necrosis and cell lysis will lead to an increase in the mobility of water in the tissue microenvironment. This increase in water diffusion translates to an increase in measured ADC as illustrated in Figure 4.1. Thus, by quantifying tumour ADC before and after therapy, tumour response or lack of response to treatment can be determined. Clinically, visual analysis of DWI and ADC data
tends to be used. Quantitative evaluations are however increasingly used, with ADC for the evaluation of early treatment response shown to detect microstructural changes that precede changes in tumour size [49,50] as illustrated in Figure 4.2. Most commonly, the mean or median ADC (ADC<sub>mean</sub>, ADC<sub>median</sub>) are estimated for tumours, within segmentations conducted on structural MRI or b-value images. More advanced metrics have however been proposed, e.g. the total diffusion volume (tDV) for assessment of global disease burden [51].

DWI for the purpose of whole body tumour assessments is increasingly employed and has shown promising results in applications such as multiple myeloma and lymphoma [49,50]. Although promising, several development needs have been highlighted to enable more widespread clinical usage, including standardisation of imaging protocols, increased accuracy of ADC estimations, signal intensity normalisation and standardisation, and development of automated tumour segmentation methods [29,52,53]. Some of these aspects were covered in Paper III and IV of this thesis.

![Graph showing therapy induced changes in tumour ADC mean](image)

Figure 4.2. Therapy induced changes in tumour ADC<sub>mean</sub> can be seen to precede changes in tumour volume for a subject with lymphoma. The patient received CAR T-cell therapy, with pre- and post-therapy PET/MRI scans performed for the purpose of therapy response assessment.

**FDG positron emission tomography**

The increased rate of glycolysis of tumour cells in most types of cancers is considered a hallmark of malignancy [16] and has made FDG PET a widely used technique for tumour imaging. As with diffusivity, tumour metabolism is potentially also altered prior to changes being detected on structural imaging [54]. FDG PET imaging, most commonly in combination with CT, has been extensively used in tumour assessments for a range of cancers, at all stages of routine cancer care; detection, staging, therapy response assessment and disease monitoring over time [55,56].

The SUV is the most commonly used metric for quantitative evaluation of tumours in FDG PET images. Due to its semi-quantitative nature it has limited
value at diagnosis, but protocol standardisation efforts [57] have made it fea-
sible to use the SUV for therapy response assessments and disease monitoring
over time [58]. Several types of SUV metrics are in use as exemplified in Fig-
ure 4.3. In addition to these, it has been suggested that SUV reproducibility
can be further improved by relating the lesion uptake to the uptake in a refer-
ence region to create tumour-to-background ratios [59]. Common reference
regions include the liver or blood, with the mediastinal blood pool used for
normalisation in Paper II of this thesis. In addition to the SUV, quantitative
FDG PET metrics related to tumour volume are in use. The metabolic tumour
volume (MTV) measures the volume of FDG-avid disease for a lesion, while
for the total lesion glycolysis (TLG) the amount of FDG uptake in each lesion
is also taken into account (TLG = MTV × SUV_mean, within the lesion). By add-
ing the MTV and TLG of all FDG-avid lesions in a whole-body dataset, a
measure of total tumour burden can be obtained. The total (or global) MTV
and TLG require all lesions in a dataset to be accurately delineated and hence
suffer from the same limitations as the SUV_mean with regards to defining tu-
mour borders in FDG PET images (see Figure 4.3).

Figure 4.3. Different types of SUV metrics commonly in use. Left: axial FDG PET
image of a lymphoma patient with one distinct tumour showing high FDG uptake.
Right: Zoomed in axial FDG PET image of the tumour with SUV metrics highlighted.
SUV_mean is the mean SUV within a segmented tumour volume. Due to image noise
and limited spatial resolution of PET, tumour borders are not well defined; SUV_mean
can hence suffer from poor reproducibility. The maximum SUV (SUV_max, the maxi-
mum voxel value in a segmented tumour volume) has higher reproducibility, mainly
as the maximum voxel value within an imaged tumour is typically invariant with re-
spect to small shifts in the tumour segmentation. The SUV_max, however, has the dis-
advantage of potentially being more biased and sensitive to image noise. A measure
less sensitive to image noise, while also not being reliant on an accurate tumour de-
lineation, is the peak SUV (SUV_peak). This is calculated from the mean SUV within a
fixed sized region of interest (ROI) containing the highest uptake region in a tumour.
4.1.1 Imaging in lymphoma

Hybrid FDG PET/CT whole-body imaging is currently the recommended imaging modality for clinical care staging, treatment response assessment, and disease surveillance in lymphoma [60–62]. The majority of lymphomas are FDG-avid, with higher uptake in aggressive lymphomas compared to indolent subtypes [61]. Common FDG distributions of DLBCL are exemplified in Figure 4.4, with large variations in whole body tumour burden and tumour sizes typically seen. The addition of CT provides structural imaging for the purpose of tumour size measurements and anatomical localisation of FDG uptake. Although FDG PET/CT is the standard for imaging in lymphoma, FDG PET/MRI has shown comparable results for diagnosis [63,64]. DWI has also been described as promising for assessing treatment response, with pre-therapy lymphoma tumours exhibiting low ADC values [65] and increased ADC values detected after successful therapy [66–68].

Current guidelines for therapy response assessment in lymphoma (see Chapter 4.3) do not necessitate quantitative analysis, although it is stressed that quantitative parameters for assessing tumour burden and therapy response would be advantageous [61]. In recent years, there is mounting evidence that semi-quantitative and quantitative metrics have a role in lymphoma treatment prediction. The total tumour burden measured by FDG PET (MTV and/or TLG) have for example shown predictive value in DLBCL after standard first-line treatment [69] and in r/r disease treated with immunotherapy [70]. The latter was also found in Paper I of this thesis.

Figure 4.4. Example of different whole body tumour burdens and distributions for five patients with DLBCL. Tumour segmentations (blue) overlaid of FDG PET maximum intensity projections (MIPs) for each patient. (This research was originally published in JNM. Barrington et al. Time to prepare for risk adaptation in lymphoma by standardizing measurement of metabolic tumor burden. J Nucl Med. 2019;60:1096-1102 [71]. © SNMMI)
4.2 System level imaging

In addition to the study of tumour structure and function, whole-body cancer imaging gives the opportunity for non-invasive system level assessments of normal organs and tissues [72].

In cancer, there is evidence that successful cancer immunotherapy is accompanied by a complex systemic response from lymphoid organs [73]. Activation of T-cells has been linked to metabolic changes [74], provoking studies of FDG PET whole-body imaging to assess metabolic patterns of lymphoid organs during immunotherapy. For immune checkpoint inhibitor therapy in melanoma, increased glucose metabolism of the bone marrow and spleen have for example been linked to favourable responses [75,76]. In paper I of this thesis, the potential of lymphoid tissue metabolism having a predictive role in CAR T-cell therapy was explored, with longer PFS seen for LBCL patients with higher pre-therapy bone marrow FDG uptake (Figure 4.5).

Figure 4.5. Bone marrow FDG uptake as measured with PET/MRI before (t0) and after (t1) CAR T-cell therapy. Longer PFS and OS observed for patients with higher pre-therapy bone marrow SUVmean. In general, bone marrow SUVmean decreased post-therapy. (From: “Whole body FDG PET/MR for progression free and overall survival prediction in patients with relapsed/refractory large B-cell lymphomas undergoing CAR T-cell therapy”. Sjöholm et al. Cancer Imaging 2022. 22:76. Licence: CC BY 4.0)

4.3 Therapy response assessment

Tumour size measurements are routinely performed for evaluating cancer response to therapy. The standard approach for structural imaging is to use the modified response evaluation criteria in solid tumors (RECIST v. 1.1), stipulating set rules for tumour size measurements [43]. RECIST has the advantages of being simple to implement and well validated. Problems have
however been highlighted with this approach, including initial increase in tumour size not always being related to treatment failure. Furthermore, RECIST incorporates limited guidelines on how to report on integrated PET/CT and PET/MRI multi-parametric data, and provides an insufficient mapping of tumour heterogeneity.

Several initiatives for improved tumour response assessments in specific settings have been published, e.g. the PET Response Criteria in Solid Tumors (PERCIST) [77] and the immune RECIST (iRECIST) [78]. PERCIST was developed as an aid for reporting quantitative FDG PET imaging findings, recommending usage of the percentage change in SUVpeak for response assessment, while iRECIST can be used for reporting on treatment response during immunotherapy. The four possible tumour response outcomes used in RECIST (complete response, partial response, stable disease, and progressive disease) is in iRECIST expanded with a fifth possible response, pseudoprogressive disease. Pseudoprogressive disease can be described as a response after initial disease progression, and has been observed in a subset of patients undergoing immunotherapy. As such, usage of iRECIST allows for continued treatment after progression.

For lymphoma, the 2014 Lugano classification details current therapy response criteria [62]. According to these guidelines, PET/CT imaging is the standard for staging and response assessment in FDG-avid lymphoma. For interim and end-of-treatment response assessment, the Deauville 5-point scale is used for tumour grading, in which the tumour FDG uptake is compared to the FDG uptake in the mediastinum and liver. Depending on the Deauville score and a comparison with baseline, the therapy response is assigned into one of four possible response categories: complete metabolic response (CMR), partial metabolic response (PMR), no metabolic response (NMR) or progressive metabolic disease (PMD).

4.3.1 Clinical endpoints

To assess the safety and efficacy of new cancer therapies, clinical endpoints are used. The purpose of the clinical endpoint is to objectively measure the benefit of the therapy in terms of a patient’s e.g. symptoms, function and survival. A common clinical endpoint is the Overall Survival (OS), defined as the time from randomisation to death. For a more rapid assessment of clinical benefit, surrogate endpoints are in use. A surrogate endpoint can be defined as ‘a biomarker intended to substitute for a clinical endpoint’ [7]. Examples include the Progression Free Survival (PFS), defined as the time from randomisation to first evidence of disease progression or death, and imaging biomarkers as described below.
4.3.2 Imaging biomarkers

A biomarker is a characteristic that objectively measures healthy or pathological processes within the body. It can stem from e.g. genomic and molecular profiling, but also from medical imaging. In the context of therapy response assessment in cancer, imaging biomarkers include e.g. the tumour, node, metastasis (TNM) staging system, objective response in RECIST and FDG PET SUV$_{max}$. A prognostic biomarker informs on patient overall treatment outcome, without taking type of therapy into account. A predictive biomarker on the other hand, informs on therapy effect and can be used for patient selection or early response assessment after therapy [79].

To become robust tools in clinical research and routine clinical care, biomarker validation in terms of technical and clinical aspects, and cost effectiveness, is needed [8]. To facilitate the usage of quantitative imaging biomarkers, the Quantitative Imaging Biomarkers Alliance (QIBA) was initiated in 2007 [80]. The initiative aims to reduce measurement errors of quantitative imaging biomarkers by developing, standardising and optimising acquisition protocols and image analyses, as well as display and reporting methods. A number of QIBA committees have been setup, with published profiles outlining quantitative accuracy achievable for e.g. FDG PET/CT for measuring response to cancer therapy [81] and usage of DWI in multi-centre clinical trials [53].

One important aspect of describing the technical performance of a quantitative imaging biomarker include measurements of precision [82]. Precision relates to the closeness of agreement between repeated measurements and types of precision include repeatability and reproducibility. Repeatability corresponds to repeated measurements over a short period of time with identical or close to identical conditions. This is often assessed using test-retest exams as performed for whole-body DWI in Paper IV of this thesis. Reproducibility corresponds to replicate measurements with conditions that vary between measurements. Scanning the same healthy volunteer on different MR imaging systems as performed in Paper IV of this thesis is an example of an assessment of reproducibility. Precision is particularly important when a quantitative biomarker is used for measuring change over time. In a therapy response assessment scenario, it is for example vital to know whether a measured change relates to a true biological change or is due to measurement variability.
Medical image processing

Since the introduction of digital imaging systems, computer-aided image processing has become an inherent part of medical imaging. To take advantage of the enormous amount of information obtained from modern imaging systems, the demand for computer-aided image processing is on the increase. In the context of cancer imaging, this includes automated workflows for segmentation of tumours, as well as detailed characterisations of tumours.

This chapter focuses on two major components of image processing: registration and segmentation. Image registration techniques can be used to integrate images into spatial alignment. Applications include e.g. aligning images from multiple subjects in a cohort study, aligning images acquired longitudinally for the same subject to study change, or combining images acquired with different modalities for the same subject. This chapter outlines the techniques used for whole-body image registration in this thesis. In cancer imaging, the aim is to distinguish the tumour from healthy tissue and extract meaningful measures. Commonly extracted metrics were described in Chapter 4, while methods for tumour segmentation are described in this chapter.

5.1 Image registration

Image registration is the process of anatomically aligning multiple images into a common reference space. In image registration, the fixed or reference image corresponds to the image that remains unchanged and the moving image to the image that is geometrically transformed using the fixed image as a reference. A transformation is produced, which allows for moving and mapping between the fixed and moving image spaces.

One approach for registration is to iteratively search for the optimal geometric transformation between fixed and moving spaces. The registration methods used in this work are based on voxel intensity, for which a similarity metric between the fixed and moving images is optimised as illustrated in Figure 5.1. The search strategy is defined by an optimiser, which specifies the strategy for finding the optimal deformation field. As a last step, an interpolator is used to resample the voxel intensities of the moving image into the fixed
image coordinate space according to the geometric transformation found. Linear, nearest neighbour and cubic interpolations are typically used for interpolation in medical imaging.

![Diagram of image registration algorithm](image)

**Figure 5.1.** Basic components of an image registration algorithm. Study A corresponds to the moving image, while Study B corresponds to the fixed or reference image. (From “Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132”. Brock et al. Medical Physics. 2017 44:7 [83]. Reprinted with permission from Medical Physics.)

**Geometric transformation**

Geometric transformation is the model that defines the type of transformation allowed in the registration process. Registration methods can also be characterised by the type of geometric transformations allowed in the registration process. One commonly used type of geometric transformation is affine transformations. Such transformations can be either rigid or non-rigid. Rigid affine transformations permit rotation and translation, while preserving shape and size of the original image. Non-rigid affine transformations additionally permit shearing and scaling of the moving image. Affine transformations tend to be used for registration of structures such as the skeleton or brain, and for pre-registration prior to a more advanced geometric transformation.

In most instances of medical image registration, a deformable transformation model is however desirable (see review in e.g. [84]). Compared to affine transformation, deformable methods are spatially varying and as such can be locally applied, only affecting voxels in close proximity of each other. Several transformation parametrisation methods for deformable registration exists, of which free form deformation is one example [85]. Control points are here defined on a grid placed on top of the image. The deformation is performed by moving the control points and as such deforming the grid (and with it, the image). Splines are commonly used to produce a continuous deformation field.

The free form deformation method is an example of a parametric model, i.e. it can be described by a set of explicit parameters. Non-parametric models also exist, which are typically represented by deformation fields and in which
each point in the moving image can be displaced arbitrarily. An advantage of many parametric methods is that the parametrisation itself introduces some degree of regularity (smoothness) to the obtained deformation. With non-parametric methods, such regularity assumptions must typically be explicitly included in the registration process to obtain reasonable transformations and avoid e.g. fold-over artefacts in image space.

**Objective function**
The objective function, or similarity metric as indicated in Figure 5.1, calculates the similarity between overlapping regions of the fixed and moving images, i.e. measures how well the images are aligned. The choice of similarity measure depends on the type of input images. For registration of images from the same modality, it is common to use sum of squared differences or normalised cross correlation. For multimodality image registrations, the mutual information (MI) metric is preferable, as this measure does not make any underlying assumptions of the image intensity distributions.

Parameter constraints, or regularisation, are often included in the objective function for non-parametric models, to impose smoothness and avoid unrealistic deformations. The diffusion regularisation is commonly used, which penalises large first order derivatives [86].

**Optimiser**
The aim of the optimiser is to find the optimal set of transformation parameters that gives the best image alignment in terms of the similarity metric in use, most commonly iteratively. Continuous and discrete optimisation methods are typically used. Common continuous optimisers in medical imaging include gradient descent, and its modification adaptive stochastic gradient descent [87]. For non-parametric models, the space of possible deformation vectors can be discretised and finding the optimal transformation viewed as a discrete labelling problem. Examples of discrete optimisers include graph-based methods [86].

As a part of the optimisation strategy, a multi-resolution strategy can be employed. Stepwise smoothing or downsampling of the images are performed, giving a pyramid of images at different resolutions. Image registration is first performed at the lowest resolution, with the transformed moving image passed to the next resolution level. As a result, coarse details of the image are first registered, followed by finer details as the image resolution is increased.

5.1.1 Imiomics
The image analysis concept Imiomics is currently being developed [88]. Within Imiomics, a cohort of 3D whole-body images can be registered to a reference space. As a result of this transformation, a voxel-to-voxel correspondence between input images is obtained, making it possible to perform
voxel-wise statistical analysis (Figure 5.2). As such, Imiomics allows for automated voxel-wise analysis of the whole body. Potential usage areas include anomaly detection and characterisation, longitudinal analysis (e.g. therapy evaluation), group comparisons and correlation between imaging data and non-imaging markers.

Figure 5.2. In Imiomics, image registration of 3D whole-body structural water-fat MR images (here visualised by two-dimensional, 2D, coronal slices) can be used for voxel-wise alignment of cohort data to a reference space. After registration, there is a voxel-to-voxel correspondence between images, enabling statistical analysis on a voxel-level. Integration of non-imaging biomarkers for the purpose of correlation analysis is also possible.

The first method for whole-body image registration in Imiomics was outlined by Strand et al [88] and registers whole-body water-fat images using a three-step pipeline according to Figure 5.3. The skeleton is first registered using a combination of affine and B-spline deformable registrations, followed by registration of lean and adipose tissues using B-spline deformable registration. Bone, lean tissue and adipose tissue are considered as separate compartments, with different elasticity constraints.
Modifications of the original registration method was used in Papers II and IV of this thesis. In paper II, a three-step pipeline was used, but lean and adipose tissues were registered using a non-parametric method, with graph-cut based optimisation [89]. The three-step pipeline used in Paper II was however prone to include fold-over artefacts, giving non-physiological displacements in registered images, in particular for the spine. To improve on this, a single-pass registration approach was developed with a compositive update rule to further regularise the produced transformation [90]. The single-pass approach utilises a regularisation weights map in which different elasticity constraints can be included for lean and adipose tissue. Compared to the initial Imiomics setup, bone was not segmented and used as a first step in the registration. Instead of using the bone registration for initial rough alignment, body masks were used for this purpose in Paper IV. In both Paper II and IV, a multi-resolution strategy for image registration was used. Example results obtained from registering the normal atlas water-fat MRI data to the water-fat MRI data of two lymphoma patients are illustrated in Figure 5.4.

Also utilised in Papers II and IV are the intrinsic co-registration of simultaneously acquired MR and PET image data obtained on a PET/MRI system. The same transformation as obtained from the registration of water-fat MR images can be used to align simultaneously acquired DWI and PET images to reference space.
Figure 5.4. Example of whole-body image registration results using the pipeline described in Paper IV. Normal atlas data was registered the patient reference space. Water MRI coronal slices are shown for a male patient (a) with corresponding registered male normal atlas (b), and a female patient (c) with corresponding registered female normal atlas (d). For the atlases, mean water MRI images are shown. Good registration results were achieved for both patients. The male patient had a large tumour burden (MTV=337 ml), while the MR images of the female patient were affected by hip implants.

5.2 Tumour segmentation

Tumour segmentation is the process of recognising tumours in an image dataset, followed by delineation of the spatial extent of the tumours with the aim to extract semi-quantitative or quantitative information as illustrated in Figure 5.5. Regardless of imaging modality, segmentation has traditionally been carried out manually by an expert reader (e.g. a radiologist). Manual segmentation is still considered the gold standard in many applications, but has the disadvantages of being time-consuming and exhibits large intra- and inter-observer variabilities. In the context of personalised medicine, the demand for quantitative tumour evaluations are likely to increase, e.g. by total tumour burden assessments as described in Chapter 4.1.1 or assessments of intra- and inter-patient tumour heterogeneity [91]. To be feasible, timely and reproducible segmentations are needed, making the manual approach unfavourable.
Figure 5.5. Overview of the standard workflow for tumour quantification in medical imaging, including image acquisition of patient data, tumour segmentation and extraction of quantitative information. In this example, a coronal slice of a patient with lymphoma imaged with FDG PET is shown, with one visible tumour segmented in red. The segmentation was automatically performed using the probabilistic atlas based approach of Paper II. After segmentation, the tumour can be characterised, either by standard metrics such as the SUV or volume, or by more advanced features for tumour heterogeneity assessments.

For evaluation of segmentation method accuracy, it is necessary to know the true boundary of the tumour. Ground truth can be obtained from histopathologic samples, but is often not available. Instead, the most common method to evaluate a segmentation method’s performance is by comparison with manual segmentation as a reference standard. The Dice similarity coefficient is a region-based metric commonly used for this purpose, in which the spatial overlap between the segmented volume \(V_1\) and the reference standard volume \(V_2\) is measured according to Equation 5.1 [92]. Segmentation methods commonly also give rise to false negative (i.e. a tumour was not segmented) and false positive (i.e. a tumour was mistakenly segmented) volumes.

\[
D_{\text{ice}}(V_1, V_2) = \frac{|V_1 \cap V_2|}{|V_1| + |V_2|} \tag{5.1}
\]

A number of semi-automated and automated tumour segmentation methods have been proposed. Automated methods have traditionally been developed for a specific body site and/or for a specific modality. Automated brain tumour segmentation is for example well-studied [93]. One class of methods that has been successfully used for segmentation in a number of applications is multi-atlas based approaches [94]. These methods essentially treat a segmentation as an image registration problem, in which an atlas of manually pre-labelled regions and the image dataset of interest are spatially aligned to facilitate transfer of labelled regions. Probabilistic atlas-based segmentation can be viewed as a special case of multi-atlas segmentation. This is commonly used in brain imaging applications for voxel-based morphometry [95], but also for tumour segmentation [93]. In this thesis, probabilistic atlas-based segmentation was used for whole body tumour segmentation (Paper II) and as input in
a deep neural network based approach for whole body tumour segmentation (Paper IV).

In recent years, deep neural networks, in particular convolutional neural networks (CNNs), have increasingly been used for automated tumour segmentation [96]. Deep neural networks are composed of multiple hidden layers in which each layer consists of several neurons connected to each other as illustrated in Figure 5.6. Training of a neural network is commonly performed using a supervised learning process. The training here refers to the process of minimising an objective function (also called loss function) by updating the network’s parameters (weights and bias terms) to achieve improved segmentation accuracy according to reference standard segmentations. Dice loss is one of the most commonly used loss functions for image segmentation [97]. Practically, k-fold cross validation is performed to test the efficacy of the model and reduce the overall bias.

In a CNN the hidden layers consists of convolutions, i.e. filters of small extent that can extract feature maps from input images. Since the last decade, the most commonly used CNN for automated image segmentation has been the U-Net [98]. Significant effort is currently ongoing to improve performance of the network by adjusting its architecture [99]. As an example, the original U-Net was implemented for 2D images, but has since been extended to 3D images [100] as employed in Paper IV of this thesis.

![Figure 5.6. A simplified illustration of a deep neural network with input (orange), two hidden (blue) and output (green) layers. In the context of tumour segmentation, the input layer corresponds to images of the tumour to be segmented, while the output layer computes the final output of the network in the form of a binary segmentation mask. The arrows indicate weights and bias terms that are tuned during training. The neurons of each hidden layer include activation functions that introduce non-linear relationships in the model.](image)

5.2.1 Whole body applications

For whole body oncology applications, methods for fully automated tumour segmentation have increased rapidly for FDG PET-based applications; in gen-
eral making use of supervised CNN based models [101]. A recent grand challenge spurred the development of many new approaches for whole body tumour segmentation using a large open-source database of FDG PET/CT cancer images [102]. In general, a Dice score of 60-70% have been achieved for these models [101], with a Dice score of 70% being described as a minimum acceptance level for this type of task [103].

For DWI, automated segmentation approaches have been published for single organ applications such as for brain tumours [104], but publications on automated whole body tumour segmentation are lacking. However, initiatives for data collection and analysis with the aim to perform automated tumour segmentation using machine learning have been described [105,106].

In addition to the network architecture, a number of other factors affect the performance of a machine learning model, e.g. factors related to image pre- and post-processing [100], and information on context [96]. One approach of the latter is to provide additional contextual information as input. As an example, publications on automated tumour segmentation in whole-body PET have focused on adding anatomical contextual information to the network [107,108]. As a variation of this, contextual awareness in the form of healthy tissue deviations are assessed in Paper IV of this thesis (Figure 5.7).

Figure 5.7. Example of automatically segmented tumours for one lymphoma patient. Coronal MIPs of (a) normalised b=900 s/mm² data and (b) healthy tissue deviations (t-map). Automatically predicted tumours obtained from training of a 3D U-Net model with (c) 2 input channels (WF and b=900 s/mm²) and d) 3 input channels (WF, b=900 s/mm² and t-map). The segmentations are illustrated in green for true positive voxels, in red for false negative voxels and in blue for false positive voxels. The Dice score between reference standard and predicted segmentations were 11% for 2 input channels and 27% for 3 input channels. Automated tumour segmentation for this patient was difficult due to several tumours having relatively low signal intensity on b=900 s/mm² images, and a high signal intensity in normal bone. MIPs are shown in inverted grey scale.
Contributions

Multi-parametric whole-body imaging data of healthy volunteers and cancer patients have been used in this thesis. The former for creating multi-parametric normal atlases and the latter to explore possibilities of voxel-wise analyses in a whole-body cancer imaging context.

Healthy volunteer data was obtained from two sources; retrospectively from the Epihealth study [109] and prospectively as a part of this thesis (Table 6.1). For the prospective study, data collection was performed between January 2019 and February 2020. Data was acquired on 1.5T MRI and 3T PET/MRI systems (DWI atlas, Table 6.1), with DWI sequences adhering to current guidelines [29] and incorporating RPG distortion correction.

Cancer patient data was obtained from a clinical trial [110] in which patients with r/r LBCL received CAR T-cell therapy and were scanned pre- and post-therapy on a 3T PET/MRI system (CAR T-cell, Table 6.1). Importantly, patients were scanned on the same PET/MRI system as the healthy volunteers.

Table 6.1. An overview of source data, including subject demographics and basic scan parameters.

<table>
<thead>
<tr>
<th>Epihealth</th>
<th>DWI atlas</th>
<th>CAR T-cell</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scanner type</strong></td>
<td>PET/MRI</td>
<td>MRI</td>
</tr>
<tr>
<td><strong>Field strength</strong></td>
<td>3T</td>
<td>1.5T</td>
</tr>
<tr>
<td><strong>MRI</strong></td>
<td>water-fat</td>
<td>water-fat</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>DWI*</td>
</tr>
<tr>
<td><strong>PET</strong></td>
<td>FDG</td>
<td>-</td>
</tr>
<tr>
<td><strong>Subject type</strong></td>
<td>HV</td>
<td>HV</td>
</tr>
<tr>
<td><strong>n, all</strong></td>
<td>89</td>
<td>39</td>
</tr>
<tr>
<td><strong>n, female</strong></td>
<td>46</td>
<td>18</td>
</tr>
<tr>
<td><strong>Age, median (range)</strong></td>
<td>68 (51-77)</td>
<td>41 (25-77)</td>
</tr>
<tr>
<td><strong>BMI, median (range)</strong></td>
<td>23.4 (18.0-35.5)</td>
<td>25.4 (19.5-34.4)</td>
</tr>
</tbody>
</table>

HV, Healthy volunteer; age [years]; BMI [kg/m²]
*b=0, 50, 400, 900 s/mm², **b=50, 400, 900 s/mm²
6.1 Paper I

Whole body FDG PET/MR for progression free and overall survival prediction in patients with relapsed/refractory large B-cell lymphomas undergoing CAR T-cell therapy

*Equal contributions as first authors, **equal contributions as senior authors

Aim
To find semi-quantitative and quantitative PET/MRI metrics of tumour and non-malignant lymphoid tissue for early prediction of PFS and OS in patients with r/r LBCL undergoing CAR T-cell therapy.

Materials and Methods
Sixteen patients (median age 63 years, range 37-71 years, 9 females) with r/r LBCL underwent CD19-targeted CAR T-cell therapy, with whole-body FDG PET/MR imaging pre-therapy and three weeks post-therapy. Tumours and normal lymphoid tissue (bone marrow and spleen) were manually segmented. Semi-quantitative and quantitative metrics were extracted, and the metric-wise rate of change (Δ) between post-therapy (t₁) and pre-therapy (t₀) calculated: Δ = metric (t₁)/metric (t₀). Tumour evaluations were performed with total tumour burden metrics (MTV, TLG and structural volume, Vₜₐₜₜ) and lesion-wise metrics (SUVₘₑₐₜ, SUVₘₐₓ, ADCₘₑₜ and structural volume, V). For calculation of Δ, target lesions for SUVₘₑₜ, SUVₘₐₓ, ADCₘₑₜ and V were identified at t₀ and t₁ for each patient, indicated as Δ(t₀) and Δ(t₁), respectively.

Univariate Cox proportional hazards regression analysis tested the relationship between extracted metrics, and PFS and OS. In addition to extracted metrics, predictor variables also included the Lugano classification at t₁. Survival curves were produced using Kaplan-Meier analysis for variables found statistically significant during the univariate analysis, with the difference in PFS and OS between subgroups assessed using a log-rank test. Correction for multiple comparisons was made with the false discovery rate (FDR). Uncorrected and corrected p-values <0.05 were considered statistically significant.

Results
During the follow-up time, progression was seen for 15 patients and death occurred for 13 patients. The median PFS was 3.9 months, while the median OS was 3.2 months from time of CAR T-cell therapy. The median follow-up time for survivors was 42.6 months (range 36.0-48.2 months).
Pre-therapy volume metrics were significantly associated with PFS: MTV (hazard ratio, HR=1.47, p<0.001), TLG (HR=1.02, p=0.0054), Vₜₐₜₜ
Association with OS for pre-therapy metrics was seen for MTV (HR=1.27, p=0.046), TLG (HR=1.03, p<0.001) and V (HR=1.90, p=0.0043). These findings remained significant after correcting for multiple comparisons (FDR<0.05), with the exception of the pre-therapy MTV association with OS (FDR=0.19). Larger tumour volumes were associated with shorter survival. MTV (≤39.5 ml) separated patients into two groups according to PFS (p=0.027, Figure 6.1a).

For rate of change metrics, ΔMTV (HR=1.71, p=0.0093), ΔTLG (HR=1.46, p=0.022) and ΔV_total (HR=1.95, p=0.024) were associated with PFS. ΔV_total was also associated with OS (HR=1.74, p=0.031). For these metrics, a larger increase post-therapy was associated with shorter survival. For ΔSUV_max, ΔADC_mean and ΔV, target lesion identification post-therapy gave significant associations with PFS and/or OS. Δ(t1)SUV_max was associated with PFS (HR=2.34, p=0.011) and Δ(t1)V was associated with OS (HR=1.09, p=0.038). Larger increases in tumour metabolism and tumour volume were associated with shorter survival. Δ(t1)ADC_mean was associated with both PFS (HR=0.83, p=0.0091) and OS (HR=0.99, p=0.040), with a larger decrease in ADC_mean post-therapy associated with shorter survival. Tumour rate of change metrics were not statistically significant after correction for multiple comparisons (FDR>0.05). For the Kaplan-Meier analysis, ΔMTV (≤1.35) and ΔTLG (≤1.35) separated patients into two groups according to PFS (p<0.001) (Figure 6.1d,e), while Δ(t1)ADC_mean (≤0.92) separated patients into two groups according to PFS (p<0.001) and OS (p=0.0054) (Figure 6.1f,g).

Pre-therapy bone marrow SUV_mean was associated with PFS (HR=0.24, p=0.0044) and OS (HR=0.28, p=0.014), with a higher bone marrow FDG uptake pre-therapy corresponding to longer survival. After correcting for multiple comparisons, association with PFS remained statistically significant (FDR=0.023). Bone marrow FDG uptake was able to separate patients into two groups according to PFS (p<0.001) and OS (p=0.0083) (Figure 6.1b,c). Spleen and rate of change bone marrow metrics were not associated with PFS or OS.

Conclusion
In r/r LBCL patients undergoing CAR T-cell therapy total metabolic tumour burden, tumour ADC_mean and FDG uptake in bone marrow unaffected by tumour infiltration are possible PET/MR parameters for prediction of PFS and OS. The findings from this explorative study suggest that PET/MRI can be a feasible imaging modality for CAR T-cell therapy evaluation in LBCL, and that a combination of FDG PET/MRI-derived imaging metrics may be useful for therapy outcome prediction.
Figure 6.1. Kaplan-Meier survival curves and log-rank p-values for extracted tumour and bone marrow metrics. The median for each metric was used for thresholding. Pre-therapy results showing MTV (a) and bone marrow SUV\_mean (b,c). Post-therapy rate of change results showing $\Delta$MTV (d), $\Delta$TLG (e), and $\Delta(t_1)$ADC\_mean (f, g). (From: “Whole body FDG PET/MR for progression free and overall survival prediction in patients with relapsed/refractory large B-cell lymphomas undergoing CAR T-cell therapy”. Sjöholm et al. Cancer Imaging 2022. 22:76. Licence: CC BY 4.0)
6.2 Paper II

A whole-body FDG PET/MR atlas for multi-parametric voxel-based analysis

*Equal contributions as first authors.

Aim

To develop a multimodality whole-body atlas of functional FDG PET and anatomical water-fat MRI data of healthy adult volunteers for the purpose of multi-parametric voxel-based analysis. As a proof of concept, a secondary purpose of the study was to perform automated voxel-wise anomaly detection using the atlas.

Materials and Methods

Eighty-nine healthy volunteers (median age 68 years, range 51-77 years, 46 females) were investigated with whole-body FDG PET/MRI. Conservative normal atlas inclusion criteria were used to ensure that all atlas entries represented high-quality scans with normal physiological values. As such, subjects were excluded due to suspected pathological findings (n=35), and image quality and protocol non-adherence issues (n=16). A multi-parametric normal atlas was created, with whole-body water-fat MR image registration used to transform PET/MR images of healthy volunteers into male and female reference spaces (see flowchart in Figure 6.2).

Figure 6.2. Flowchart showing input and output image data (squares) and major processes (circles) involved when creating the multi-parametric atlas. (From “A whole-body FDG PET/MR atlas for multi-parametric voxel-based analysis”. Sjöholm et al. Scientific reports 2019. 9:6158. Licence: CC BY 4.0)
Subjects with poor registration results were excluded (n=6). Specific aspects of incorporating PET data into the normal atlas were considered, including FDG PET intensity normalisation to the mediastinal blood pool and bias assessment to detect subjects with outlier FDG PET distributions. The latter lead to subject exclusion due to increased leg muscle FDG uptake (n=3).

Quantitative measurements of fat content, \( \text{SUV}_{\text{mean}} \) and tissue volume were performed to assess the adequacy of atlas content. As proof of concept, voxel-wise statistical comparisons were performed between the atlas and (i) two subjects with suspected malignant disease and (ii) one subject with elevated liver fat infiltration.

Results
An atlas containing quantitative parametric maps of water and fat fraction MRI, local tissue volume and FDG PET uptake was created with separate entries for male (n=12) and female (n=15) healthy volunteers. Manual segmentation in the male and female reference spaces confirmed that the atlas contained adequate physiological and anatomical values for the majority of tissue FDG uptake, tissue fat content and organ size. A quantitative bias was however noted for adipose tissue, for which the male and female FF, and male SUV measurements, showed statistically significant differences between native and reference space. This is likely due to abnormal stretching of fat, occasionally obtained in the registration process.

The atlas was applied in two anomaly detection tasks as proof of concept. The first task automatically detected anomalies in one male and one female subject with suspected malignant disease using FDG PET data (Figure 6.3). After tumour size and FDG uptake restrictions, 19 out of 20 tumours were automatically detected. One false negative lesion corresponded to a larger lesion in the male thoracic vertebrae. Problems noted included failure to separate close lying tumours and false positive tumours being detected from physiological uptake in the brain, bladder, heart and intestines. The second task successfully detected abnormal liver fat infiltration in one subject using fat fraction data.

Conclusion
A fat and water fraction MRI, local tissue volume and FDG PET normal atlas was created using data acquired on a hybrid PET/MRI scanner and a newly introduced image registration method. As proof of concept, two anomaly detection tasks were performed using the atlas. The first task successfully detected non-normal FDG uptake for two subjects with suspected tumour disease. The second task automatically found abnormalities in liver fat infiltration.
Figure 6.3. Results from a proof of concept anomaly detection task performed for a female volunteer with suspected lymphoma. Representative coronal slices are shown of registered water fraction and normalised FDG uptake images, and of the automatically detected and manually segmented tumours. Segmented tumours are displayed in red and overlaid on the water fraction image. (From “A whole-body FDG PET/MR atlas for multi-parametric voxel-based analysis”. Sjöholm et al. Scientific reports 2019. 9:6158. Licence: CC BY 4.0)

6.3 Paper III

Improved geometric accuracy of whole body diffusion weighted imaging at 1.5T and 3T using reverse polarity gradients

Aim
To assess the performance of the RPG method for correction of susceptibility-induced geometric distortion of whole-body DWI healthy volunteer data acquired at 1.5T and 3T. For clinical applicability, to assess whether $b=50 \text{ s/mm}^2$ ($b_{50}$) DWI data can be used for RPG correction instead of standard $b=0 \text{ s/mm}^2$ ($b_0$) DWI data.

Materials and Methods
Whole-body DWI and structural MRI data of adult healthy volunteers were acquired at 1.5T ($n=20$, median age 44 years, range 25-77 years) and 3T ($n=20$)
median age 46 years, range 27-76 years). DWI and structural water-fat MRI data were collected station-wise using five or six axial stations. DWI data was acquired using a diffusion-weighted spin echo EPI sequence with STIR fat suppression in free breathing. Four b-values were acquired: b=0, 50, 400 and 900 s/mm². For the standard DWI sequence, PE was in the anterior-posterior (AP) direction. For RPG distortion correction, a reverse DWI sequence was acquired with PE in the posterior-anterior (PA) direction. DWI data was distortion corrected using the RPG method based on b0 and b50 DWI acquisitions. An illustration of the steps involved in the distortion correction is shown in Figure 6.4.

Figure 6.4. Illustration of the steps involved in the RPG geometric distortion correction using b0 data. The CMTK tools epiunwarp, reformatx and imagemath were used as indicated [111]. Epiunwarp calculates the 3D deformation field map from b0 AP and PA input images. The resulting deformation map then corrects all b-value images using the reformatx and imagemath tools. Reformatx applies the deformation map to each b-value image, while a voxel-wise multiplication with the Jacobian of the deformation is carried out using imagemath, correcting for signal pile-up. NC, non-corrected; DC-b0, distortion correction using b0 data; b400, b=400 s/mm²; b900, b=900s/mm². (From “Improved geometric accuracy of whole body diffusion weighted imaging at 1.5T and 3T using reverse polarity gradients”. Sjöholm et al. Scientific reports 2022. 12:11605. Licence: CC BY 4.0)

The RPG method’s performance was assessed in terms of (1) station-wise geometrical alignment between DWI and structural MRI data using the MI similarity metric, (2) effect on healthy tissue ADC values, and (3) visual evaluation of image quality. The visual evaluation was performed on whole-body non-corrected and distortion-corrected b50 images in two steps assessing (a)
spine misalignment at station boundaries on sagittal images and (b) overall axial image quality as scored independently by three radiologists.

Results
Larger geometric distortions were observed at 3T compared to 1.5T, with the distortion correction also seen to give larger geometrical improvements at 3T. Examples of the effect of the distortion correction at 3T are shown in Figure 6.5.

Figure 6.5. The effect of distortion correction and typical registration artefacts at 3T. Axial b50 images (a), axial b50 images overlaid on structural images (b) and sagittal b50 images (c). In each column (left to right) the non-corrected (NC) AP, NC PA, b0-based distortion corrected (DC-b0) AP and b50-based distortion corrected (DC-b50) AP images are displayed. Improvement in geometrical alignment is exemplified by zoomed axial b50 images in yellow colour scale overlaid on corresponding structural images (b) and by sagittal b50 images in inverted grey scale (c). Misalignments in NC AP images are indicated with arrows (b,c). For sagittal images, station borders are marked with a dashed line. Kidney registration artefacts are marked by black arrows (a). (From “Improved geometric accuracy of whole body diffusion weighted imaging at 1.5T and 3T using reverse polarity gradients”. Sjöholm et al. Scientific reports 2022. 12:11605. Licence: CC BY 4.0)

MI between DWI and structural data increased for all stations when distortion correction was applied (Wilcoxon signed-rank tests, p<0.05). Small numerical differences between non-corrected and distortion corrected healthy tissue ADC values were measured.

Visually, the distortion correction improved spine alignment at station borders. Without distortion correction, 14 subjects (70%) had at least one station boundary with spine misalignment at 1.5T and all subjects (n=20) had at least one spine misalignment at 3T. After distortion correction, the spine was
aligned at all station boundaries for all subjects at 1.5T. At 3T, the distortion correction decreased the number of misalignments, but five subjects had one remaining misalignment after correction. Radiologists scored the overall image quality of distortion-corrected data as worse for the majority of subjects at 1.5T (mean n=15 subjects, 73%) and 3T (mean n=13 subjects, 63%), mainly attributed to minor artefacts introduced by the registration (see example in Figure 6.5).

Results for the b0- and b50-based distortion corrections were comparable for all assessments.

Conclusion
Geometric distortions were present in the majority of acquired data, with larger distortions observed at 3T compared to 1.5T. An improved geometric accuracy was obtained when applying the RPG distortion correction on whole-body DWI data acquired at 1.5T and 3T. The b0- and b50-based distortion corrections had an equal performance.

6.4 Paper IV
A whole-body diffusion MRI normal atlas: development, evaluation and initial use
Therese Sjöholm, Sambit Tarai, Filip Malmberg, Robin Strand, Alexander Korenyushkin, Gunilla Enblad, Håkan Ahlström and Joel Kullberg. Manuscript

Aim
To create, evaluate and employ a normal atlas of whole-body DWI and ADC of healthy volunteers scanned at 1.5T and 3T. The atlas was created using deformable image registration and evaluated by establishing whole-body ADC values of healthy tissue, with test-retest ADC measurements, comparison of ADC across field strength and assessments of the effect of age and gender on ADC. We further employ the normal atlas in an automated tumour segmentation task, together with a deep learning approach, to investigate whether healthy tissue deviations could be advantageous in this type of task.

Materials and Methods
Forty-five adult healthy volunteers were scanned at 1.5T (n=38) and/or 3T (n=29) using a whole-body MRI protocol including water-fat MRI and DWI. Twenty-three subjects were scanned at both field strengths and test-retest imaging was performed for five subjects per scanner.

Using deformable image registration, sex- and BMI-stratified normal atlases of high b-value DWI and ADC were developed. ADC was measured using manual volume of interest (VOI) segmentations for ten healthy tissues:
parietal white matter, cerebellar white matter, liver, spleen, kidneys, psoas muscle, vertebral body, pelvic bone, femur and thigh muscle. Test-retest ADC\textsubscript{mean} was evaluated using the percentage repeatability coefficient (%RC) \cite{53}, while the effects of gender and scanner on ADC\textsubscript{mean} were studied with group-wise t-tests and Bland-Altman analyses, respectively. Multiple linear regression was used to test if age and gender significantly predicted ADC\textsubscript{mean} for the VOI-based analysis, and Pearson correlation was used to assess the voxel-wise correlation between age and ADC for registered whole-body images in male and female reference spaces.

Lastly, a framework for using the normal atlas in an automated tumour segmentation task was setup. For this purpose, lymphoma PET/MR imaging data and reference standard segmentations from Paper I were utilised. A 3D U-Net architecture with and without information about healthy tissue deviations were setup, with healthy tissue deviations measured by voxel-wise statistical comparisons between patient and normal atlas DWI data. The predicted segmentations were evaluated in terms of Dice score between reference standard and predicted tumour segmentations, and lesion-wise sensitivity and precision.

\textbf{Results}

Age- and BMI-stratified whole-body normal atlases containing parametric maps of healthy tissue DWI and ADC were created at 1.5T and 3T (Figure 6.6).

Healthy tissue ADC\textsubscript{mean} measurements showed a large variation in %RC depending on tissue assessed, ranging from <10\% for the brain to 48\% at 1.5T (liver and muscle) and 70\% at 3T (spleen). Bland-Altman plots showed large scanner differences, with statistically significant differences between ADC\textsubscript{mean} measured at 1.5T and 3T for all assessed tissues except the kidneys, pelvic bone and vertebral bodies.

Gender differences in ADC\textsubscript{mean} were measured for liver, thigh and psoas muscle, femur and vertebral bodies at 1.5T, but only for thigh muscle at 3T. At 1.5T, age was negatively associated with ADC\textsubscript{mean} for the liver, vertebral body and femur. At both field strengths, a positive association between age and ADC\textsubscript{mean} was obtained for parietal white matter. These findings were confirmed in the voxel-wise whole body correlation analysis at 1.5T for female subjects, for which negative correlations between ADC and age were observed for e.g. liver, vertebral bodies and femur, and positive correlations between ADC and age were observed for brain tissue.

Adding information about healthy tissue deviations in an automated tumour segmentation task increased the median Dice score numerically from 40\% to 45\%. The lesion-wise sensitivity and precision also increased; from 33\% to 42\% for sensitivity and from 65\% to 69\% for precision.
Figure 6.6. Example normal atlas images for male BMI $\geq 25$ kg/m$^2$ healthy volunteers scanned at 1.5T (n=11). Coronal WF, $b=900$ s/mm$^2$ and ADC images are shown. The top row corresponds to mean images, while the bottom row corresponds to the percentage coefficient of variation (%CV).

**Conclusion**

In this work, sex- and age-stratified whole-body DWI and ADC normal atlases were created at 1.5T and 3T. The atlases were used to study voxel-wise correlations between healthy tissue ADC and age across the whole body, confirming results from a manual segmentation approach. A deep learning based framework for automated tumour segmentation was setup. Statistical deviations between lymphoma subject and normal atlas DWI were shown to numerically improve Dice score, sensitivity and precision of this task.
Discussion

The studies included in this thesis were conducted to improve the current state of cancer imaging and image analysis methods in whole-body MRI and PET/MRI. In a broad perspective, the whole-body image analysis methods applied and developed during this work can be viewed as a contribution to the research field ‘whole person health’. As opposed to traditional reduction-based research in which physiological systems within the body are studied as separate processes, this concept relates to investigations of the whole body by e.g. studying multiple physiological systems in healthy and diseased states, or during therapeutic interventions. Whole person research has been highlighted as a key objective in the US National Center for Complementary and Integrative Health’s strategic plan for 2021-2025 [112].

In this chapter, work carried out as part of this thesis is further discussed with regards to the imaging contrasts used, diffusion imaging standardisation, quantitative therapy evaluation, image analysis automation and limitations. The chapter ends with a summary of the key findings of this thesis.

7.1 Imaging contrasts

Medical imaging consists of a number of imaging contrasts that are important tools for detection and characterisation, as well as longitudinal monitoring, of cancers. In this thesis, whole-body imaging data composed of water-fat MRI, diffusion MRI, and FDG PET were used. Although these represent techniques widely used in cancer imaging applications, the field is constantly evolving and a large number of additional imaging contrasts are either already available or being developed.

Water-fat imaging was used for structural imaging in this thesis. Besides producing images for tumour size measurements, these images are particularly useful for detection and characterisation of bone metastases. Other structural T1-weighted MR imaging techniques exist however (e.g. 3D turbo spin echo), but these do not have the advantage of water and fat separation. The water and fat separation was also a prerequisite for the whole-body image registration method used in this thesis.
In this work, the radiotracer FDG was used for PET imaging. This is the most commonly used cancer-imaging radiotracer, targeting the increased metabolism seen for a majority of malignant tumours. One disadvantage is however that FDG is non-specific, meaning benign FDG uptake can also be obtained, most commonly due to inflammatory processes [113]. Inflammatory changes and increased tumour metabolism are hence indistinguishable in FDG PET images, which can pose problems when using FDG for early therapy response assessments, as in Paper I of this thesis. A large number of non-FDG PET radiotracers are available and have shown utility in specific settings, e.g. usage of $^{68}$Ga-PSMA in prostate cancer and $^{68}$Ga-DOTA for neuroendocrine tumour imaging. Of particular interest to this thesis is in vivo CAR T-cell imaging. Novel radiotracers are under development that aim to visualise CAR T-cell activity after infusion, facilitating in vivo monitoring of the therapy [114]. If successfully translated into humans, these radiotracers could aid in the assessment of efficacy and toxicity of new CAR T-cell therapies, and potentially give indications for pre-selecting patients who may benefit from CAR T-cell therapy.

For whole-body DWI, current guidelines recommend imaging with a minimum of two b-values in the ranges 50-100 s/mm$^2$ and 800-1000 s/mm$^2$, and usage of a mono-exponential calculation of the ADC (Equation 3.8). Imaging at b=50-100 s/mm$^2$ means perfusion-related signal is minimised, while the upper b-value limit is set to provide good contrast of hyper-cellular tumours with sufficient SNR. Imaging at even higher b-values would be advantageous in terms of tumour contrast, but is prohibitive due to a reduction in SNR. Computational methods have however been developed with the aim to increase tumour contrast while preserving SNR, including the computed DWI [115] and voxel-wise computed DWI [116] as used in Paper IV of this thesis. In addition, there is the potential to image with an extended number of low b-values to study tissue perfusion. This concept is used in intravoxel incoherent motion imaging, with oncological imaging suggested as a promising usage area [117]. This contrast has the potential to also become available for whole-body imaging [118].

New imaging techniques are also being introduced which could change the status of recommended imaging contrasts in oncology applications. In the context of whole-body imaging, total body PET was recently launched. With this technology, it is possible to image the entire body (axial FOV 70-200 cm depending on system), with an increased sensitivity of $\times10$-40 [119]. This offers the potential to e.g. track small quantities of radiolabelled cells in vivo and, in addition to static FDG PET measures, study dynamic processes across the whole body. Photon-counting CT is another novel imaging technique recently introduced clinically [120], in which energy-resolved CT data can be acquired. Compared to conventional CT, increased soft tissue contrast and spatial resolution at a reduced ionising radiation dose can be obtained.
7.2 Diffusion imaging standardisation

A prerequisite for the widespread use of any medical imaging technique is a standardised approach for data collection and analysis. Whole-body DWI is increasingly used in oncology applications [121], but its acceptance as a quantitative imaging biomarker has been slow due to e.g. lack of optimised and standardised imaging protocols and heterogeneity in ADC measurement methodology. Guidelines on protocol optimisation and standardisation have however been published in recent years [29,121], with the aim to improve reproducibility across scanners and sites.

Image artefact reduction has been highlighted as important for ADC usage as a primary endpoint biomarker [29]. One challenge in whole-body DWI protocol optimisation is magnetic field non-uniformities obtained when covering a large anatomical area. As observed in Paper III of this thesis, susceptibility-induced spatial distortion can be large, in particular at 3T. Paper III further showed that this type of distortion can be substantially reduced using the RPG method as a post-processing step. This method was selected as it has been described as robust to large distortions [35], as typically seen in whole-body DWI. Other schemes for distortion correction however exist including field mapping [33] and registration approaches [122], but were not assessed in the current work. Slice-specific shimming has also shown promising results for reducing geometric misalignments, but currently requires long acquisition times [123].

To move from qualitative to quantitative DWI assessments, knowledge about measurement precision and accuracy is needed. The QIBA profile for usage of DWI in multi-centre oncology clinical trials was published to facilitate appropriate use of quantitative DWI. The current profile includes precision measurements for brain, liver, prostate and breast applications, with %RC of 11-47% reported depending on organ [53]. Whole-body DWI is not included due to lack of sufficient literature on test-retest data. Paper IV of this thesis adds information on this topic by analysing test-retest data of several healthy tissues across the whole body at 1.5T and 3T. The measurements performed showed that ADC repeatability varied largely depending on scanner and imaged tissue. This is likely due to multiple factors, including variations in distortion depending on anatomy, signal reception with different coils (e.g. head and surface coils) and tissue differences in blood perfusion. Reduced SNR of high b-value images also affects the ADC calculation. In this regard, it is possible that whole-body test-retest imaging of tumours instead of healthy tissues would have given an improved repeatability, as tumours tend to maintain a higher SNR in high b-value images.

For the QIBA profile claim to be valid, longitudinal scans must be performed on the same scanner, using the same scan protocol and the same analysis methods. The need for this statement was highlighted in Paper IV of this
thesis, which showed that ADC varied significantly for healthy tissue assessments performed on two different scanners.

7.3 Quantitative therapy evaluation

To take advantage of the personalised medicine approach that modern cancer therapies can offer, the basic tumour response assessment criteria currently in use, consisting of a few one-dimensional measures, might not be sufficient. As most malignancies are heterogeneous, as much as possible of the extent of the disease must instead be mapped (e.g. total tumour burden and intra- and inter-lesion heterogeneity) to allow for therapy selection and response evaluation. In this context, and to take full advantage of all information available from modern medical imaging systems, advanced quantitative tumour measurements could be of great value.

Basic quantitative measurements such as tumour size from structural images and extraction of SUV metrics from FDG PET images are to some extent already in clinical use, while other measures have shown promising results (e.g. ADC, MTV and TLG). In addition, increasing amounts of metrics for tumour heterogeneity quantification have been proposed, often via a radiomics approach [124]. Within radiomics, quantitative features related to e.g. intensity, shape and texture can be extracted from segmented tumours. Several studies have demonstrated the predictive and diagnostic ability of radiomics features in different types of cancers using various imaging modalities [125]. Worth noting, however, is that standardisation is of great importance for reliable extraction of quantitative metrics and correct interpretation of results. For example, it has been shown that features extracted from PET image data can vary significantly depending on factors such as acquisition protocol, reconstruction method and segmentation method [126].

Advanced quantitative evaluations can be split into region- and voxel-based approaches. In a region-based approach, quantitative metrics are extracted from ROI or VOI segmentations, often covering the tumour as a whole. In a voxel-based approach, of which Imiomics is one example, parametric response mapping is possible. This gives the opportunity to visualise different spatial zones within a tumour on the voxel-level and has been used for e.g. tumour response mapping using DW image data [127] and prediction of local recurrence sub-regions using FDG PET image data [128].

Using multimodality imaging, joint assessment of contrasts from different imaging modalities is possible, in many cases providing complementary information. In this thesis, hybrid PET/MR imaging in lymphoma was used in Paper I, with a number of PET and MRI metrics showing promising results for prediction of PFS and OS during CAR T-cell therapy. Due to the limited number of patients, the current study could however not investigate the complementary value of extracted metrics.
7.4 Image analysis automation

Qualitative assessments of medical images have been used for a long time, with humans having an extensive capacity for visual analysis. Visual analysis is however subjective and time-consuming, and as such prohibits routine usage of quantitative evaluations. Image analysis has successfully been used for automation in many applications, with the methods used and developed in this thesis contributing to the field of automated whole-body image analysis.

Whole-body normal atlases
In this thesis, deformable image registration of water-fat MRI data was used to produce whole-body normal atlases. The whole-body registration method has evolved over time, primarily by work presented by Ekström et al [89,90]. As an example, total body water-fat imaging was required for the registration pipeline used in Paper II. This was however not feasible for oncology applications, in which whole-body imaging (eyes to thighs, or similar) tends to be used. As such, the registration pipeline used in Paper IV was developed to handle these types of scans. Most likely, the whole-body registration approach will continue to evolve, and be adapted to specific settings and patient cohorts. Large differences in BMI were present in both healthy volunteer and CAR T-cell patient cohorts in Paper IV. As a result, it was not sufficient to use different elasticity constraints for lean and adipose tissues in the registration, as proposed by Strand et al [88]. Usage of an inside SAT mask was in this case needed for improved registration results of subcutaneous adipose tissue, and the normal atlases were also stratified according to BMI. For cohorts with narrow BMI ranges, an inside SAT mask would likely not be needed.

By including subjects with wide BMI and age ranges and a cancer patient cohort, it was observed that the registration method could handle non-normal deviations in images. This was exemplified in Figure 5.4 for patients with high tumour burden and metal implants. The datasets used in this thesis were however of limited size and testing with larger cohorts are needed for a deeper understanding of the capacity and limitations of the registration method.

The whole-body normal atlases were in this thesis employed in healthy tissue assessments in terms of voxel-wise correlation analyses and in automated anomaly detection and segmentation tasks. Further potential usage areas include longitudinal monitoring of cancer, in which the registration can aid in tracking individual tumours over time, and studies of the tumour and its microenvironment using a parametric response mapping approach. As the registration method is based on water-fat MR images, it has the advantage of being PET tracer independent, with possibilities to perform studies of non-FDG PET/MRI patient cohorts. In addition, as highlighted in Chapter 4.2, whole-body imaging gives the opportunity to study systemic changes and interactions across the whole body. Paper I showed that an increased pre-therapy bone marrow FDG uptake gave longer PFS and OS after CAR T-therapy in a cohort...
of r/r LBCL patients, potentially indicating that bone marrow hyperactivity is a part of a systemic immune response.

**Automated tumour segmentation**

Tumour segmentation is a critical component of extraction of quantitative imaging biomarkers. Repeatable and accurate methods are needed for comparing data longitudinally and across sites. Automated tumour segmentation approaches have mainly been developed for single organ or limited FOV applications, although whole-body tumour segmentation is making rapid progress primarily using deep learning based models.

An added component of complexity encountered for whole-body imaging data, compared to single organ or limited FOV applications, is that tumour appearances (e.g. shape, texture, signal intensity) often depend heavily on the organ involved. The strategy used for tumour segmentation must hence be flexible in this sense. As an example, tumour FDG uptake in one organ might equal healthy FDG uptake in another organ, meaning threshold-based methods do not work satisfactorily. FDG PET and DWI also suffer from poor spatial resolution and noise. Low resolution means a decreased contrast between the object of interest and the background is obtained, making it hard to determine where to place the tumour border. The fuzzy borders of tumours in FDG PET data is a well-studied problem. As mentioned in Chapter 4.1, SUV$_{\text{max}}$ and SUV$_{\text{peak}}$ are often used to avoid defining the exact tumour border.

Two methods for automated whole-body tumour segmentation were used in this thesis, a probabilistic atlas based approach in Papers II and IV, and a deep learning based approach in Paper IV. Regardless of method, signal intensity normalisation techniques were utilised for both FDG PET and high b-value DW images. This was needed to harmonise inter-subject signal intensities. Static FDG PET imaging can provide semi-quantitative information using the SUV. There are however a number of potential problems linked to the SUV, including both physical and biological sources of error, that lead to inter-subject variability. By normalising to a region inside the body, as is commonly performed in both brain and body FDG PET applications, variability can be reduced. In Paper II, whole-body FDG PET data was normalised to the FDG uptake in the mediastinal blood pool by automated transfer of manually segmented aortic ROIs in reference space. Compared to FDG PET, DWI has the added problem of arbitrary units for b-value images, and for whole-body imaging, the technique suffers from both intra- and inter-subject signal intensity variations. Different methods for DWI signal intensity normalisation was assessed in this thesis, including Z-score, robust Z-score and upper quartile normalisation [129], as well as histogram based methods [130] and normalisation to a reference region. It was however not possible to achieve the same level of inter-subject signal intensity harmonisation as for FDG PET data. Further work in this regard is needed.
Probabilistic atlas based segmentation has most extensively been used in brain imaging applications. This thesis however makes initial attempts to extend this concept to whole-body imaging. For FDG PET, a proof of concept study was performed in Paper II with promising results. It was noted, however, that false positive tumour segmentations occurred for normal physiological FDG uptake and in the presence of image registration errors. Extending the probabilistic atlas based segmentation approach to whole-body DWI in Paper IV, proved difficult due to the larger intra- and inter-subject signal intensity variations seen for this imaging contrast, and the presence of image artefacts. After signal intensity normalisation, a large number of false positive voxels were in general segmented, indicating that differences in signal intensities prevailed after normalisation.

In paper IV, a deep learning based approach was used for automated tumour segmentation of whole-body DWI data, with median Dice scores of 40-45% achieved. Compared to similar studies performed for whole-body FDG PET data, for which Dice scores of 60-70% are typically reported [101], this represents inferior results. To some extent, this can be explained by the patient cohort used, containing a limited number of scans (n=40) and many small tumours with low contrast on high b-value DWI. The larger intra- and inter-subject signal intensity variations seen for whole-body DWI, and increased presence of image artefacts, likely also affected the segmentation accuracy negatively.

Adding contextual information to the deep learning model, from the probabilistic atlas based segmentation approach, gave improvements in Dice score, sensitivity and specificity. The deep learning model was hence able to extract meaningful information from the probabilistic atlas based segmentation, even though a large number of false positive voxels was obtained using this approach on its own. By increasing the quality of the contextual information provided, it is likely that the segmentation performance could improve further. Factors that could improve the probabilistic atlas based segmentation include developments in intra- and inter-subject signal intensity normalisation and in the whole-body registration framework.

Future work, to further improve the tumour segmentation performance, naturally includes expanding the amount of training data, but also taking advantage of the multimodal nature of the data by exploring synergistic effects of FDG PET and DWI data as input in a deep learning model. In general, small datasets is a major problem for deep learning applications in medical imaging. Current methods rely on manually annotated data for training and validation, which is a limiting factor for creating large datasets. Multi-centre collaborations are in general needed to provide sufficient training data. Between-scanner ADC and DWI signal intensity was in this study however shown to be large, and age and gender differences were shown to exist for a subset of organs. For multi-centre studies, these factors might prove problematic. Im-
Improvements in signal intensity normalisation techniques are needed, and potentially also age- and gender-matched studies. Another important challenge is acceptance of deep learning methods by the radiology community. As it is difficult to intuitively understand how deep learning works, there is a lack of transparency, and as such challenging for medical doctors to base their clinical decisions on. In the context of tumour segmentation, this is not primarily related to the Dice score achieved, but to ensure that tumours are detected if present.

7.5 Limitations

A limitation highlighted in all studies included in this thesis, is the relatively small number of subjects available: to create normal atlases, to perform automated tumour segmentations and to predict survival after CAR T-cell therapy. As a result, the work presented should be viewed as preliminary and further studies are needed to fully understand the possibilities of the presented methods. Limited data availability is a common problem in medical research. In recent years, however, the radiology community has initiated a move towards data sharing in the form of open-access medical imaging repositories, as well as code sharing. Advantages are many, including datasets being more widely used, diverse datasets being combined, increased transparency and increased opportunities for collaboration.

For voxel-wise whole-body statistical analysis a large number of statistical tests are performed, giving a multiple comparison problem. This was to some extent considered in the current work by using spatial smoothing, strict p-values and clustering. Numerous correction methods however exist, but were not used in the current work. For whole-body analysis, permutation-based approaches have shown promising results [131], and could be applied in future studies.

Lastly, distortion corrected DWI data for lymphoma patients was not available in the studies performed. This would have been preferable as the geometric distortion of DWI and ADC data can give large discrepancies between diffusion and structural imaging data. To study the effect of distortion correction on tumours would have been advantageous in Paper III. The automated tumour segmentation approach of Paper IV would most likely also have benefitted from geometrically aligned structural and diffusion MRI datasets. Although technical improvements such as parallel imaging and multi-channel surface coils mean that whole-body PET/MRI with DWI can be performed in a reasonable time frame (approximately 45 min), the study time prohibits the addition of further imaging sequences. For future studies it might however be clinically feasible to include distortion correction, as faster sequences for e.g. RPG distortion correction are becoming available.
7.6 Conclusion

The contributions of this thesis to the current state of cancer imaging and image analysis methods in whole-body MRI and PET/MRI include:

I. The usage of multi-parametric whole-body FDG PET/MRI for predictive assessments of CAR T-cell therapy outcome in aggressive large B-cell lymphoma. The total metabolic tumour burden, tumour \( \text{ADC}_{\text{mean}} \) and FDG uptake in bone marrow unaffected by tumour infiltration were possible PET/MR parameters for prediction of PFS and OS.

II. The usage of deformable image registration to produce a multi-parametric whole-body PET/MRI atlas of healthy volunteers. Initial proof of atlas usage for automated whole body assessments of healthy tissues and tumours were provided.

III. The assessment of the RPG method for geometric distortion correction of whole-body DW imaging data. In a cohort of healthy volunteers, the geometric accuracy between structural and diffusion images increased with small effects on healthy tissue ADC values.

IV. The usage of deformable image registration to produce whole-body diffusion MRI atlases. The atlases were used to study voxel-wise correlations between healthy tissue ADC and age across the whole body, confirming results from a manual segmentation approach. A deep learning based framework for automated tumour segmentation was setup. Statistical deviations between lymphoma subject and normal atlas DWI were shown to numerically improve Dice score, sensitivity and precision of this task.

Eftersom majoriteten av cancersjukdomar är potentiellt systemiska kan bildgivande helkroppsundersökningar vara användbara både vid diagnostikfallet och vid uppföljande terapiutvärderingar. Bildalstrande metoder som positronemissionstomografi (PET) och magnetresonanstomografi (MRT) har genomgått snabb utveckling så att det nu är möjligt att skapa högupplösta tredimensionella helkroppsbilder i syfte att upptäcka primärtumörer och metastaser hos cancerpatienter samt att följa upp dessa efter behandling. PET och MRT möjliggör simultan insamling av morfologiska och funktionella bilddata. Typiska kännetecken för cancer (t.ex. cellproliferation, hypoxi, ökad metabolism, apoptos and angiogenes) kan med dessa system följas över tid och på så sätt potentiellt generera detaljerade tumörkaracteriseringar. Den enorma mängd data som skapas av moderna bildgivande system innebär dock en utmaning vad gäller tillvaratagande av all insamlad bildinformation. Traditionell visuell bedömning av den omfattande bildinformationen är subjektiv, kräver erfarenhet och mycket stor noggrannhet, och är därmed tidsödande. För att mer effektivt ta tillvara på den bilddata som finns tillgänglig behövs avancerade och automatiserade metoder för bildanalys som kompletterar den visuella bedömningen.

Målsättningen med denna avhandling är att utveckla avancerade metoder för bildtagning och bildanalys i syfte att karakterisera tumörer med hjälp av helkroppsbilder från MRT och kombinerad PET/MRT.

**Delarbete I**

I detta arbete utvärderades semi-kvantitativa och kvantitativa mått hämtade från helkroppundersökningar utförda med PET/MRT av patienter med aggressivt storskaligt B-cellslymfom. Syfte var att predicera överlevnad efter behandling med CAR T-celler. Patienterna undersöckes med PET/MRT före och tre veckor efter terapi. Bilddata analyserades baserat på manuella segment-
teringar av samtliga tumörer och lymfvävnad. Arbetet visade att total metabolisk tumörvolym, apparent diffusion coefficient (ADC) i tumör och upptag av fludeoxyglucose (FDG) i icke-malign benmärg är potentiella PET/MRT-markörer för prediktion av överlevnad.

**Delarbete II**
Genom att använda deformabel bildregistrering utvecklades i detta arbete en multimodal normalatlas i helkropp av vuxna friska frivilliga, med syftet att utföra multiparametrisk voxelvis analys. Bilddata bestod av vatten-fett MRT, lokal vävnadsvolym och FDG PET insamlad på ett PET/MRT-system med fältstyrka 3T. Som konceptvalidering användes denna utvecklade normalatlas för voxelvisa statistiska jämförelser med PET/MRT-bilder på frivilliga forskningspersoner med sjukdomsfynd. För två frivilliga med misstänkta tumörsjukdomar visades att voxelvis analyser av FDG PET-upptag med hjälp av normalatlasen automatiskt segmenterade misstänkt tumörbörda. För en av de frivilliga forskningspersonerna visade voxelvis analyser av fetthalt, erhållt från vatten-fett MRT, en förhöjd nivå av leverfett.

**Delarbete III**
Geometriska förvrängningar är vanligt förekommande i diffusionsviktad MRT (DWI), vilket medför att dessa bilder skiljer sig georeferentiskt från andra simultant insamlade bildkontraster (t.ex. morfologisk MRT eller PET). Denna diskrepans kan vara problematisk i både visuell/manuell och automatisk bildanalys. Detta arbete utvärderade en existerande metod för geometrisk korrektion av DWI, kallad reverse polarity gradient (RPG), för applikation på helkroppsdatal. Helkroppssundersökningar med DWI utfördes på vuxna friska frivilliga med en 1.5T MRT- och en 3T PET/MRT-utrustning. Bilddata korrigerades retrospektivt med hjälp av RPG-metoden. Arbetet kunde visa att användandet av RPG gav en ökad geometrisk noggrannhet för helkropps-DWI, särskilt för fältstyrka 3T.

**Delarbete IV**
Acknowledgements

The work presented in this thesis was performed at the Department of Surgical Sciences at Uppsala University, with financial support from the Swedish Cancer Society and Antaros Medical AB. In addition, this thesis was made possible by the support from numerous people. I am particularly grateful to:

The healthy volunteers and patients for participating in the MRI and PET/MRI scans that enabled this research. Thank you for your time and commitment.

My main supervisors, Filip Malmberg and Joel Kullberg. Filip guided me through the first part of this thesis. Thank you for your encouragement throughout these years. I have appreciated your good advice and friendly attitude. For the second half of the thesis, Joel was my main supervisor. Thank you for always having solutions to my problems, I could always count on that! Thank you also for friendship and for creating a pleasant atmosphere in the office.

Co-supervisors Robin Strand and Håkan Ahlström. Håkan, thank you for your support and for good advice especially during our work on Paper I. Robin, thank you for solid research input and your always friendly demeanour.

To all my supervisors I am grateful for facilitating a flexible work environment, enabling me to pursue a PhD while being a present parent for my two toddlers. This has meant a lot to me.

Fellow staff in the research group. Elin, thank you for friendship, good advice and nice lunches. I have truly missed having you around during these last months of the PhD journey. Simon, thank you for good company at entrance 24, for a smooth collaboration on Paper II and for answering numerous registration-related questions. Alexander, thank you for an interesting collaboration on Paper I, I learnt so much! We have been lucky to have you in the office, you are such a nice person. Tomas, thank you for taking time out from your busy schedule to score multiple diffusion scans for Paper III. Thanks to Hanna for using your MD skills to inject Buscopan. Thanks Sambit, for your valuable help on Paper IV and for teaching me how to train a deep-learning model. Thanks also to Elin, Simon, Alexander and Sambit for constructive feedback on selected parts of this thesis. To all of you, including Robin V, Taro, Jonathan, Nouman and the original Antaros gang – thank you for good company
during these years. You have all, in different ways, made my time as a PhD student more joyous.

Research nurses Anders Lundberg, Gunilla Arvidsson, Marie Åhlman, Anne-Marie Montelius and Caroline Lybeck. Thank you for finding time in your busy schedules helping me in protocol optimisation, answering numerous questions and teaching me how to operate the MRI scanners.

Mathias Engström, thank you for generously sharing your in-depth knowledge about MRI physics.

Gunilla Enblad, thank you for good collaboration on Paper I.

John Dickson who was my colleague and mentor at the University College London Hospital. Thank you for helping me find my feet in my first nuclear medicine job all those years ago, giving me a solid foundation and generously introducing me to the world of research.

John Totman, my manager and friend at the Clinical Imaging Research Center in Singapore. Thank you for all your support while living in Singapore, both at work and privately. I have now finished the PhD, just not at Karolinska as you first helped me arrange. I have also been lucky to give birth to two children during the PhD, just not in Singapore as I then hoped for. You never know what life will bring! I learnt so much from you and I hope our paths will cross again.

Mamma och pappa, tack för en trygg barndom med mycket frihet att låta tankar vandra fritt och kroppen ströva (och klättra!) vart den ville. Min syster Charlotte, tack för att du alltid finns vid min sida.

Mats, Ella och Leo, tack för all kärl under denna period. Ni har konstant påminnit mig om vad som är viktigast i livet. Jag är leden för att jag stundtals har varit väldigt tankspridd.

Mats, tack för ovillkorlig stöttning! En ny tid väntar nu och jag är tacksam och glad över att du fortsatt finns vid min sida.

Ella och Leo. Här och nu är era ledord, vilket har gjort mitt arbete med denna avhandling mycket mer njutbart. Vilken lycka att just ni är mina barn!
References


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1898

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-495136

ACTA UNIVERSITATIS
UPPSALIENSIS
UPPSALA
2023