Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer

Short title: MAchine Learning In whole Body Oncology (MALIBO)

Version: 4.0, 21 August 2019

SPONSOR: Imperial College London

FUNDERS: EME NIHR and Imperial College CRUK Centre

STUDY COORDINATION CENTRE: Imperial College London

Protocol authorised by

(Name & Role):   Prof. Andrea Rockall

                         Consultant Radiologist

                         Chief Investigator

Chief Investigator: Prof. Andrea Rockall

Co-investigators: Prof Eric Aboagye, Dr Ioannis Lavdas, Prof Daniel Rueckert, Dr Ben Glocker, Dr Xinxue Liu, Prof Deborah Ashby, Shalini Santhakumaran, Nicholas Johnson, Prof Toby Prevost, Prof Stuart Taylor, Dr Shonit Punwani, Prof Vicky Goh, Dr Dow-Mu Koh, Dr Nishat Bharwani, Dr Sandy Beare

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

Table of Contents

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

Abbreviations:

CR_UK	Cancer Research UK
CT	Computed Tomography
ICTU-Ca	Imperial Clinical Trials Unit – Section on Cancer
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NIHR	National Institute of Health Research
RT	Reading Time
TMG	Trial Management Group
TSC	Trial Steering Committee
WB DW-MRI	Whole body diffusion-weighted imaging)

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

1.0 Study summary

Title	Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer **Short title:** MAchine Learning In whole Body Oncology (MALIBO)
Study Design	This is an observational study using MRI scans from three different patient cohorts, which will be used to: 1) develop and 2) evaluate machine learning algorithms over three consecutive phases
Research Questions	**Primary research questions:** 1) Is the specificity of WB DW-MRI scans, in patients being staged for cancer, significantly improved with no subsequent loss of sensitivity when ML methods are applied? **Secondary research questions:** 1) Can the reading time (RT) and associated costs of WB DW-MRI scans be reduced when ML techniques are employed to assist experienced radiologists? 2) Can inter-observer variability be reduced by the use of ML methods in experienced or new WB MRI readers? 3) Can the application of ML methods in WB MRI increase the diagnostic accuracy delivered by less experienced radiologists?
Hypothesis and aims:	**Hypothesis:** The use of ML methods that automatically identify normal anatomical structures and subsequently detect abnormal lesions can improve the diagnostic accuracy and reduce the RT of WB DW-MRI scans of patients having cancer staging. **Aims:** To apply ML techniques to WB DW-MRI in order to significantly improve specificity with no loss of sensitivity and to improve the radiology RT, so that the technique reaches the performance of highly sensitive/specific imaging modalities in oncology (such as FDG-PET/CT)
Objectives	**Primary objective:** To compare the diagnostic accuracy of WB DW-MRI, as read by experienced readers, in patients being staged for cancer, with and without the aid of ML methods against the

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

	reference standard of full clinical diagnosis at 12-month follow up (or the appropriate reference standard of each contributing study). **Secondary objectives:** 1.To compare the RT and associated costs of WB MR scans; 2. To assess inter-observer variability; 3. To test the diagnostic accuracy of non-experienced readers.

2.0 Background

Whole body MRI, including diffusion weighted MRI (DW-MRI), is currently an active research interest in oncology imaging as a non-invasive technique for the detection of metastatic disease, as well as a potential biomarker for clinical use and drug development (3). A meta-analysis has been published which supports further development of WB MRI in clinical practice, in view of the promising sensitivities and specificities in bone metastases (4). Diffusion weighted MRI (DW-MRI) is increasingly being used in WB imaging. DW-MRI allows quantification of water diffusivity in tissues and has been found to be sensitive for detecting tumour sites in organs and bones, with visible changes in the MR signal intensity due to reduction in water diffusivity associated with the highly cellular nature of tumour tissue (5). The characteristic appearances of the bone marrow have been studied in relatively small numbers of patients without metastatic disease, and in patients with breast cancer, myeloma and prostate cancer (6-8).

Other than the slow nature of manual reads, one of the main issues when using WB DW-MRI for staging of patients with cancer is the potential number of false positives. Many 'normal' anatomical structures (such as lymph nodes) may reflect similar diffusion properties compared to pathological regions. The possibility of using computer- assisted reading or machine learning (ML) techniques has been considered in aiding interpretation of complex MRI datasets. One group evaluated the topography of whole body adipose tissue and proposed an algorithm that enables reliable and completely automatic profiles of adipose distribution from the WB dataset, reducing the examination and analysis time to less than half an hour (9). Another group has developed a parametric modelling approach for computer-aided- detection of vertebral column metastases in WB MRI (10). Machine learning (ML) techniques have previously been developed to differentiate benign (86 cases) from malignant (49 cases) in soft tissue tumours using a large MR database of multi-centre, multi-machine MR images, but without using DWI (11). Co-investigators at

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

Imperial College London have previously developed methods for organ localisation in WB DIXON MRI and accurate semantic segmentation on CT (12-16).

2.1 Rationale for the proposed study:

In order to make WB DW-MRI a useful and clinically relevant tool within the NHS, a method that could assist the radiologist both in improving diagnostic accuracy whilst reducing reading time would be beneficial to deliver better accuracy, productivity and cost-effectiveness. An important aspect in the development of diagnostic support systems is semantic understanding of input data. In case of WB DW-MRI, it is essential to 'teach' the computer to automatically detect and localise different anatomical structures, and discriminate normal and pathological appearances. A computer system that is able understand what is shown in an image can be effectively used to implement an intelligent radiology inspection tool. Such a tool would greatly support the radiologist when reading the large amount of MRI data. Guided navigation to regions of interest, automatic adjustment of organ and tissue specific visualisation parameters, and quantification of volume and extent of suspicious regions are some of the features that such a system would provide and thus, drastically reduce the time needed for an expert to perform diagnostic tasks. Previous ML methods (described in 2.2) can be adapted to WB DW-MRI to allow automatic vertebrae localization, to automatically exclude false positives in suspicious regions and to discriminate malignant from benign structures (12-16). These methods have yielded promising results for their respective tasks. They are all based on a particular concept of ML called supervised learning. In supervised learning the assumption is that some annotated training data is available that can be used to train a predictor model. Here, the annotations reflect the output value that one want to infer for new patient images. The training data can be defined as a set $T = \{(X_i , Y_i)\}$ of pairs of input data X_i, here a WB DW-MRI, and some desired output Y_i, for example a point-wise probability map that indicates the likelihood for each image point to be malignant. Using the training data, the aim of an employed learning procedure is then to estimate the conditional probability distribution $(Y|X)$. Having a good estimate of this distribution allows prediction of output Y for any new input data X. In the context of WB DW-MRI for staging, the automatically obtained predictions for a new patient image can be integrated in a radiology inspection tool, for example to automatically navigate to or highlight suspicious region.

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

2.2 Clinical studies that this proposal relates to:

The proposed research will use whole body magnetic resonance (WB MRI) data predominantly from the NIHR-funded STREAMLINE L and STREAMLINE C study. These are multi-centre prospective cohort studies that evaluate WB MRI in newly diagnosed non-small cell lung cancer (250 patients; STREAMLINE-L; ISRCTN50436483) and colorectal cancer (322 patients; STREAMLINE-C; ISRCTN43958015) (1). The studies initially defined WB MRI acquisition, quality assurance and analysis protocols applicable to daily NHS practice. The objectives of both studies are the same. The primary objective is to evaluate whether early WB MRI increases detection rate for metastases compared to standard NICE-approved diagnostic pathways. Secondary objectives include assessing: influence of WB MRI on time to and nature of first major treatment decision following definitive staging. At 12-month patient follow-up, a multidisciplinary consensus panel will define the reference standard for tumour stage considering all clinical, pathological, post mortem and imaging follow-up. Accuracy will be defined per-lesion, per-organ and per-patient.

The STREAMLINE C study started in March 2013. It had recruited 128 patients by the end of January 2015 and is currently expected to complete in October 2018. The STREAMLINE L study started in February 2013. It had recruited 135 patients by the end of January 2015. It is currently expected to complete by October 2018. The study started on 1.10.2012 and is currently expected to end on October 2018. The current estimated total recruitment September 2015 is 398 (204 for L and 194 for C). The anticipated drop-out or death rates by 1 year are 20% and 10% for the L and C trials respectively. The ISRCTN for STREAMLINE L is ISRCTN50436483 and for STREAMLINE C is ISRCTN43958015 (1).

Additional cases may be obtained from the CR_UK funded MELT study if they are considered compatible with the STREAMLINE-C and STREAMLINE-L data (Whole Body Functional and Anatomical MR: Accuracy in Staging and Treatment Response Monitoring in Adolescent Hodgkin's Lymphoma Compared to Conventional Multimodality Imaging, NCT01459224) (2). Data from the UCLH MASTER study, including cases with lymphoma and prostate cancer may also be used based on compatibility (the myeloma cases from

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

MASTER are unlikely to be used). Planned enrolment, which is nearly complete (71 patients recruited by February 2015), is 100 patients. The justification for using these datasets is that they will demonstrate additional cases of nodal disease and sclerotic bone metastases thereby ensuring a variation in the distribution of disease used to develop the ML algorithm, as the cases from STREAMLINE are likely to have more non-nodal metastatic sites, such as liver and lytic bone metastases. The purpose of the MELT study is to compare staging accuracy as well as response assessment using WB MRI with standard investigations in patients with newly diagnosed Hodgkin's lymphoma. It is a prospective observational cohort study. The primary outcome measures are: per site sensitivity and specificity of MRI for nodal and extra-nodal sites and concordance in final disease stage with the multi-modality reference standard (at staging). The reference standard for the MELT study is contemporaneous MDT with all other staging eg PET CT and CT at the time of diagnosis and initial staging, Secondary outcome measures: 1) Inter-observer agreement for MR radiologists 2) evaluation of different MRI sequences on diagnostic accuracy; simulated effect of MRI on clinical management. Planned enrolment for MELT is 55 and current recruitment estimates are that 45 patients will have undergone 2 scans each (90 WB MRI scans) by September 2014, with study completion March 2015.

3.0 Study objectives

Primary objective: To compare the diagnostic accuracy of WB DW-MRI, as read by experienced readers, in patients being staged for cancer, with and without the aid of ML methods against the reference standard of full clinical diagnosis at 12 month follow-up period.

Secondary objectives:

1. To compare the RT potential cost saving of WB MRI scans with and without ML support.
2. To assess inter-observer variability with and without ML support.
3. To test the diagnostic accuracy of experienced versus non-experienced readers with and without ML support.

Exploratory objectives (data permitting):

1. To compare intra-rater variability with and without ML support.

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

4.0 Study design

This is an observational study (study limited to working with data), using three different patient cohorts, being evaluated in series during three consecutive phases.

Phase 1: Development and optimisation of ML pipeline to automatically identify anatomical structures of interest in WB DW-MRI. For automatically labelling anatomical structures of interest, we will extend previous work that automatically segments abdominal organs from Computed Tomography (CT) data (14). More specifically, we will use a hierarchical weighting approach in which the anatomical atlases will be constructed first at subject level and then followed by atlas construction at organ level and finally at voxel level. This approach has been shown to accommodate the significant body anatomical variability across different subjects. By combining this with patch-based segmentation we will be able to accurately and robustly annotate anatomical structures of interest. In order to construct the anatomical atlases, whole body MRI data sets from 50 healthy volunteers will be used; these have been collected under a separate ethics approval (08/H0707/58). The initial ML algorithm 'A' will be produced.

Phase 2 'Training set': Develop the ML pipeline for the automatic detection and identification of cancer lesions. For this we will learn shape and appearance models that are specific to the anatomical regions identified in Phase 1. These models will allow the probabilistic interpretation of the images in terms of a generative model. Classification will be carried out using advanced ML techniques based on ensemble classifiers such as random forests (17). WB MRI scans from the STREAMLINE L and C, MELT (if compatible) and MASTER (if compatible) studies with established disease stage (main study reference standard, described above) will be used to train ML detection of metastases. We estimate that approximately 60 scans with metastases and 90 without metastases will be needed to train the ML algorithm to detect tumour sites accurately. Allocation of cases to phase 2 will be undertaken by the study statistician in order to allow appropriate cases for training and "held-back" cases for validation. At least 150 WB-MR scans will be evaluated by two expert readers over the course of Phase 2 reviewing ML output to provide active learning feedback. Depending on the final number and distribution of metastatic lesions available from the source studies, additional cases may be required to improve the training of the algorithm, potentially using a total of 300 radiology cases. Initial radiology reads

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

(approximately the first 100 reads) will focus on identifying ML errors and identification of true positive lesions according to the radiology ground truth reference standard. This information will be used to gradually refine and improve the computer algorithm to develop ML 'A+', which will be fine-tuned across the next Phase 2 radiology reads against radiology ground truth to develop ML algorithm 'B'. Final Phase 2 reads against radiology ground trutth will be used for any further refinements to complete the final ML algorithm 'C'. An analysis of per lesion sensitivity will be performed at this stage using approximately 40-50 new (held-out) patient datasets (to allow for sufficient positive cases). Sensitivity for detection of metastatic lesions by the algorithm can be evaluated using a DICE co-efficient metric, with sufficient overlap and probability threshold achieved.   The threshold will be determined as we develop the algorithm in the early stages of phase 2. If the upper 95% confidence interval (CI) of the sensitivity by algorithm 'C' is less than 80%, then further work on the algorithm will need to be undertaken prior to proceeding to Phase 3. Cases used for the sensitivity check at the end of phase 2 will not have been used for any ML training or be read by radiologists. Therefore, if the sensitivity level is not met, then these could be used to further improve the algorithm, prior to progressing to phase 3. If the sensitivity level is met, these cases could be part of the validation set in phase 3.

Phase 3 'Validation set': A 2nd set of WB MRI data relating to an estimated 193 subjects from STREAMLINE, plus additional data from the MELT and MASTER studies (on the condition that data is compatible with STREAMLINE), that have not been used for training will be read by multiple expert readers with ML'C' support. The per-patient specificity and sensitivity of WB MRI assessment, with and without ML 'C' support, will be determined using the established reference standard from the main study. An interim analysis of the first 50-70 consecutive cases will be undertaken. If the upper 95% CI of per patient specificity of WB-MRI with ML 'C' support does not reach 80% then further review of the algorithm 'C' will be required. RT will be recorded. Sub-studies will include: 1. Reads by new (non-expert) WB MRI readers; 2. Repeat reads across multiple readers (in random order and at time interval) with and without ML 'C' support to measure reading time and inter-observer variation. This will ensure parity in computer set-up between the reads, as there may have been variation in the original main study reads related to use of either PACS, Biotronics 3D platform or other software, in addition to differences in internet speeds when reads were performed on-line for the main study.

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

5.0 Eligibility and study population

No patients will be directly recruited into this study. Recruitment and scanning of patients has taken place under separate studies (contributing studies) with their own ethical approval. The following inclusion/exclusion criteria summarise the patient populations that are providing scanning data for this study (contributing studies).

Inclusion and exclusion criteria applied for Phase 1 (Healthy volunteers):

Inclusion criteria

1. Healthy volunteers were aged 18 to 100 years
2. Written, informed consent was provided

Exclusion criteria

1. Any co-existing medical illness
2. Contra-indications to MRI (e.g. patients with pacemakers, metal surgical implants and aneurism clips, patients suffering from claustrophobia)

5.2 Inclusion and exclusion criteria for Phases 2 and 3:

Inclusion criteria

1. Patient eligible for and consented to take part in one of the contributing studies (STREAMLINE C or L, MELT, MASTER)
2. Patient completed the study imaging assessments

Exclusion criteria

1. Patient that consented to contributing study but did not complete the scan
2. Scan could not be adequately completed

5.3 Brief description of contributing studies inclusion/exclusion criteria:

STREAMLINE L and C study inclusion criteria: histopathologically confirmed or suspected lung cancer or colorectal cancer being staged for initial treatment planning; written informed consent. Exclusion criteria include any contra-indication to MRI scanning

MELT study inclusion criteria: aged 6-18 years with participant/guardian informed consent, histologically confirmed Hodgkin's lymphoma, treated with the Euronet chemotherapy regime. Exclusion criteria: contra-indications to MRI, previous other malignancy or

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

pregnancy/nursing

MASTER study inclusion criteria: diagnosis of prostate cancer, lymphoma or myeloma.

6.0 MRI protocols

Scanning protocol applied for Phase 1:

Localiser images have been acquired and then breath-hold anatomical scans (T_1-w with DIXON and T_2-w) were acquired in four stations. The volunteer was instructed through the intercom to breath in-breath out and hold several times in each station. The table moved between stations. After this the diffusion-weighted scan (with five b-values) took place, which was free-breathing. The table again moved between the four imaging stations.

Scanning protocol for Phases 2 and 3:

Patients will be scanned according to the protocol of the source (contributing) study. All of these studies include whole body anatomical (T_1-w and T_2-w) MRI with diffusion-weighted imaging, as for the Phase 1 studies. Local protocol variations include the use of 1.5T versus 3T scanner and minor differences in the applied sequence parameters. Some of the studies also include the use of intravenous contrast (gadolinium) injection.

7.0 Machine Learning pipeline and reading process

Previous ML methods will be adapted to WB DW-MRI to allow automatic vertebrae localization, to automatically exclude false positives in suspicious regions and to discriminate malignant and benign structures (12-16). These methods are all based on a particular concept of ML called supervised learning. In supervised learning the assumption is that some annotated training data is available that can be used to train a predictor model. Here, the annotations reflect the output value that one want to infer for new patient images. The training data can be defined as a set $T = \{(X_i, Y_i)\}$ of pairs of input data X_i, here a WB DW-MRI, and some desired output Y_i, for example a point-wise probability map that indicates the likelihood for each image point to be malignant. Using the training data, the aim of an employed learning procedure is then to estimate the conditional probability distribution $(Y|X)$. Having a good estimate of this distribution allows prediction of output Y for any new input data X.

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

Phase 1

For automatically labelling anatomical structures of interest, we will extend previous work that automatically segments abdominal organs from Computed Tomography (CT) data (14). More specifically, we will use a hierarchical weighting approach in which the anatomical atlases will be constructed first at subject level and then followed by atlas construction at organ level and finally at voxel level. This approach has been shown to accommodate the significant body anatomical variability across different subjects. By combining this with patch-based segmentation we will be able to accurately and robustly annotate anatomical structures of interest. In practise, we will manually outline and annotate normal anatomical regions from the T2-w whole body images of the healthy volunteers, using a manual segmentation tool (e.g. ITK-SNAP). Labelled images will then be saved in Nifti format and fed to the ML algorithms from previous works (14) to generate ML algorithm 'A'. In order to construct the anatomical atlases, whole body MRI data sets from 50 healthy volunteers will be used.

Phase 2

Development of the ML pipeline for the automatic detection and identification of cancer lesions. For this we will learn shape and appearance models that are specific to the anatomical regions identified in Phase 1. These models will allow the probabilistic interpretation of the images in terms of a generative model. Classification will be carried out using advanced ML techniques based on ensemble classifiers such as random forests (17). → ML algorithms 'A+', 'B' and 'C'.

During Phase 2, the diagnostic radiologists will be allocated cases for review (from a pool of at least 150 cases, up to 300 cases) in order to identify lesions that have been confirmed as malignant by the source study reference standard. Each WB-MRI case will ultimately be reviewed by two radiologists during the iterative development of the machine learning algorithm. Malignant lesions will then be outlined by the radiologist on each of the available imaging sequences to establish ground truth segmentation masks of disease sites. Each outlined lesion will then be reviewed by the ML computing scientists and MR physicist. The information provided by the outlined malignant lesions will be incorporated into the ML model. Following the development of algorithm C, the next 40-50 consecutive cases from the source studies will be tested using the algorithm to evaluate the threshold

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

sensitivity (see statistics section) for detection of malignant lesions.

Phase 3

During Phase 3, the first cohort of 50-70 patients will be distributed to the diagnostic radiologists, following application of the ML support to the source data. Radiologists will not be allocated cases that they have previously read as part of the source studies, thus ensuring that all the reads in Phase 3 are blinded. Each radiologist will read the WB-MRI scans both with and without ML support, using all sequences used in the source studies. The order of allocation will be randomised by the study statistician and there will be an interval of at least 4-weeks between the two sets of radiology reads. The detection of metastatic lesions and primary lesions, incorporating ML support and radiologist's expertise, will be recorded using the pro forma used in the STREAMLINE studies, adapted to the MALIBO study. Tumour stage and reading time will be recorded. An interim analysis will be performed to evaluate the specificity of radiology reads with ML support. If the specificity target is met (see statistics section), then the remainder of the study scans will be read by the allocated radiologists. Sub-studies will be performed using additional reads to evaluate inter-reader variability and differences between inexperienced and experienced readers. An exploratory analysis may be carried out to assess intra-reader variability if at least one reader is able to perform duplicate sets of reads.

Allocation of reads to Phase 2 and Phase 3

To ensure that the algorithm developed in Phase 2 is developed and tested using similar data in Phase 3, reads were assigned to each phase utilising stratification on the following categories:

- Source of data (study)
- Site of metastases (liver, bone and nodal).
- Study site of scan origin (hospital)

8.0 Imaging assessment and patient follow-up

There is no further assessments or follow-up for Phase 1 healthy volunteers. WB DW-MRI will be assessed for the presence of disease, using an imaging volume from the brain to mid-thighs. Reads will proceed using a specific ordered viewing of sequences to follow the STREAMLIME study reads with experience and inexperienced WB MRI readers. The

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

reading time will be recorded. Assessments will be made with ML support and the results will be compared to the main study read results (without ML support) for each WB MRI study. Patient assessments and follow-up for Phases 2 and 3 will be according to the source protocol. No additional patient assessments or follow-up will be performed within the MALIBO study.

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

9.0 Sample size calculation and statistical analysis

9.1 Sample size calculation:

In a previous published meta-analysis, Wu et al 2011 reported a pooled per patient sensitivity of 88% and a pooled specificity of 86% for whole body MR with DW-MR (4). We anticipate that ML support will improve specificity by 10%, from 86% to 96% against the reference standard and will test for a difference between the WB-MR with and without ML using McNemar's test for paired proportions (18).

The comparison of specificities of WB-MR with/without ML support is summarized in the following 2 by 2 table.

Observed frequencies table:

	WB MR result with ML		
WB MR result without ML	Negative	Positive	Total No of Pairs
Negative	a	b	a+b=86%
Positive	c	d	c+d=14%
Total	a+c=96%	b+d=4%	N=100%

Sample size calculation for the testing Phase 3:

➤ Paired study design – comparing the specificities of two methods using McNemar's test[2].

➤ Since there is no background of the expected proportion of discordant pairs, we applied the following approximation:

Proportion of discordant pairs = $Specificiy_1*(1-Specifity_2)+Specificity_2*(1-Specificity_1)^2$.

The proportion of discordant pairs: $\pi= 0.96*0.14+0.86*0.04=0.169$

➤ Odds ratio of the two methods: $\psi=0.96*0.14/0.86*0.04=3.9$

➤ Type I error: one side $\alpha=0.05$ (We believe that WB MRI will be superior with ML support than without ML support)

➤ Power: $1-\beta=90\%$

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

A sample size of 141 patients with negative reference standard is needed. Since patients in the MELT study are all metastasis, we will use the data in STREAMLINE study to estimate the sample size. As the prevalence of non-metastasis is expected to be 80% based on the preliminary data of STREAMLINE study, the total sample size for Phase 3 is 177 patients from STREAMLINE. Therefore, a minimum of 177 patients from STREAMLINE study will be required to detect a difference of 10% between the specificity of WB-MR with ML support and that of WB-MR without ML support with Type I error of 5% (one-sided) and power of 90%.

Following the allocation of STREAMLINE data to Phase 2, it is expected that 193 cases from the STREAMLINE study are available for use. Amongst the 193 patients from STREAMLINE studies, 51 are expected to have metastatic disease. Based on the above, we are anticipating an improvement in specificity of 10% from 86%. Regarding sensitivity, on the assumption that the sensitivity of WB-MR with ML support is no less than that of WB-MR alone (88%), a sample size of 51 metastases will provide an expected 95% confidence interval for the sensitivity of WB-MR with ML support of 79% to 97%.

9.2 Proposed outcome measures:

Primary outcome measure will be the per patient specificity of WB MRI with ML 'C' support compared to standard radiology read (WB MRI without ML 'C' support) against the reference standard established in the main study. Specificity is defined as the proportion of patients with negative reference standard, which has been correctly classified as negative by radiologist based on WB MRI with or without ML'C' support.

Secondary outcome measures will be: 1. Reading time of WB- MRI with or without ML 'C' support with estimation of associated costs; 2. Inter-observer variability measured by the Kappa Coefficient between the experienced radiologists with and without ML support; 3. The diagnostic accuracy of WB MRI with ML 'C' support and standard radiology read (WB MRI without ML 'C' support) against the reference standard established in the main study for experienced readers versus inexperienced readers. Diagnostic accuracy will include the sensitivity, defined as the proportion of patients with positive reference standard, which has been correctly classified as positive by radiologist based on WB MRI with or

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

without ML'C' support;

9.3 Statistical analysis:

Phase 2 analysis will be undertaken as an integral aspect of the ML algorithm development and the study output from this phase will be ML algorithm 'C'.

Primary analysis, Phase 3

The per patient specificities of the two methods (with and without ML) against reference standard will be compared using McNemar's test for paired proportions.

Secondary analysis, Phase 3

Sensitivity: The per patient and per lesion sensitivity of WB MRI with and without ML support will be reported with 95% confidence intervals.

Specificity: The per lesion specificity of WB MRI with and without ML support will be reported with 95% confidence intervals.

Reading time: Reading time will be compared between WB MRI with and without the ML support adjusting for covariates and random effects of radiologist.

Inter-observer variance: Summary statistics of the proportions of concordant and discordant diagnosis between multiple experienced radiologists will be reported for both methods. Inter-observer variance will be measured by Kappa coefficient.

Diagnostic accuracy of inexperienced readers: Summary of diagnostic accuracy (sensitivity and specificity) with 95% confidence interval will be evaluated for inexperienced readers.(see 4. Study design).

Intra-observer variance: If it is possible to obtain the services of at least one radiologist for an additional session of reading, summary statistics of the proportions of concordant and discordant diagnosis between the two different read sessions (across the same subset of reads) will be reported. Intra-observer variance will be measured by Kappa coefficient.

Simple Cost-effectiveness: Cost-effectiveness for radiology reading time will be summarised.

Missing data: It is unlikely that there will be any missing data as this study will use scans and follow-up data already collected within the STREAMLINE, MELT (if compatible) and MASTER (if compatible) studies. Patients with missing data or loss to follow-up will not be

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

included in this study.

Interim analysis: Two interim analyses will be carried out during the study. The first interim analysis concerning per lesion sensitivity by the ML algorithm 'C' will be undertaken using 40 – 50 new patient datasets after Phase 2. We will require the upper 95% CI of the sensitivity no less than 80%. If this is not met, then further work on algorithm 'C' will be required. The second interim analysis will be undertaken based on the first 50-70 consecutive patient cases in Phase 3 (depending on the prevalence of metastasis in the STREAMLINE studies). An upper 95% CI of 80% for the specificity of algorithm 'C' will be required prior to progress. If this is not met, then further development of the algorithm 'C' will be required.

10.0 Regulatory issues

10.1 Ethical arrangements:

Ethical approval for all source (contributing) studies is in place. Patients from the source studies provide consent for their anonymised data to be used in future research. Ethical approval for Phase 1 is in place. Ethical approval for Phases 2 and 3 will be undertaken prior to commencement of these phases. There are no material ethical concerns related to the study with no perceived risk or benefit to individual patients but there is a significant interest in improving patient care, as indicated in s60 (1) of the Health and Social Car Act (2001). There is no perceived risk in delays to the start of the study being caused by ethical review as Phase 1 of the study can start immediately as the ethical agreement for this aspect of the study has already been obtained. If the ethics committee request further patient consent for anonymised review of WB-MR scans, there is significant lead-time prior to the start of phase 2 during which this process could take place, although this is unlikely to be required.

10.2 Research Governance:

The study is sponsored by Imperial College London. The sponsor has civil liability insurance, which covers this study in the United Kingdom.

11.0 Study management

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

The trial MRI physicist will manage the study day to day, with oversight from the CI and Imperial Clinical Trials Unit – Section on Cancer (ICTU-Cancer).

11.1 Trial Management Group:

A Trial Management Group (TMG) will be convened including the Chief Investigator, co-investigators and key collaborators, a patient representative, trial statistician, MRI physicist and trials manager. The TMG will be responsible for day-to-day conduct of the trial and operational issues.

11.2 Trial Steering Committee:

A Trial Steering Committee (TSC) will be convened including as a minimum an independent radiologist (Chair), independent computer scientist, independent statistician and a patient representative. Members of the TMG will also attend meetings of the TSC including the Chief Investigator, Lead Computer Scientist and Senior Statistician. The TSC will provide overall supervision of study conduct and progress.

12.0 Publication Policy

Information concerning the study, patent applications, processes, scientific data or other pertinent information is confidential and remains the property of the Sponsor. The investigator may use this information for the purposes of the study only. It is understood by the investigator that the Sponsor will use information developed in this clinical study and, therefore, may disclose it as required to other clinical investigators. In order to allow the use of the information derived from this clinical study, the investigator understands that he/she has an obligation to provide complete test results and all data developed during this study to the Sponsor. Verbal or written discussion of results prior to study completion and full reporting should only be undertaken with written consent from the Sponsor. Therefore all information obtained as a result of the study will be regarded as CONFIDENTIAL, at least until appropriate analysis and review by the investigator(s) are completed. The results may be published or presented by the investigator(s), but the Sponsor will be given the opportunity to review and comment on any such results for up to 1 month before any presentations or publications are produced.

13.0 Study plan flow diagram

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019
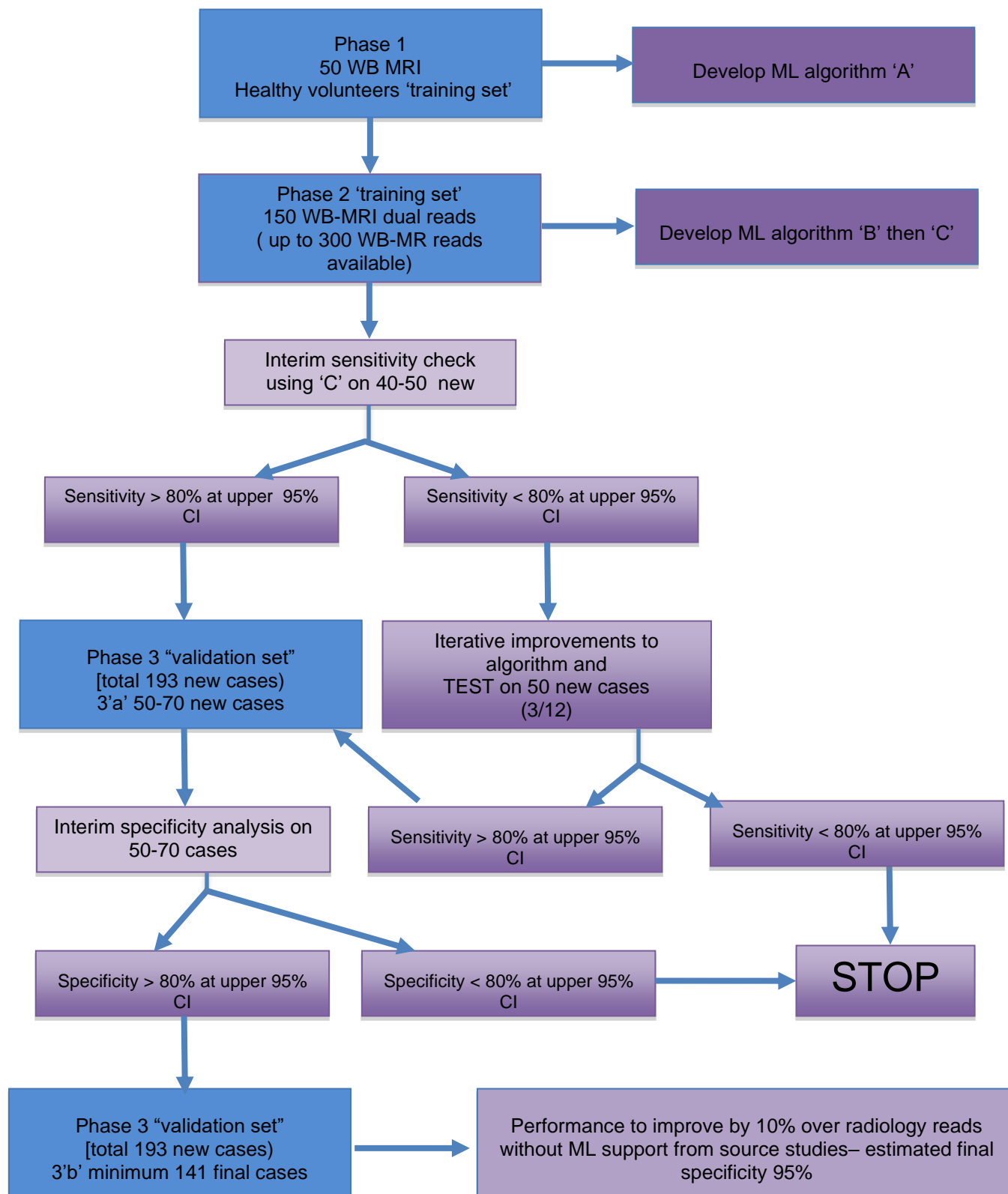
14.0 References

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

1. STREAMLINE STUDY:

http://www.nets.nihr.ac.uk/__data/assets/pdf_file/0003/81678/PRO-10-68-01.pdf

2. MELT STUDY: http://clinicaltrials.gov/show/NCT01459224

3. Padhani AR, Koh D-M, Collins DJ. Whole-Body Diffusion-weighted MR Imaging in Cancer: Current Status and Research Directions. *Radiology* 2011; **261(3)**: 700-718

4. Wu L-M, Gu H-Y, Zheng J, Xu X, Lin L-H, Deng X, Zhang W, Xu J-R. Diagnostic value of whole-body magnetic resonance imaging for bone metastases: a systematic review and meta-analysis. *J Magn Reson Imaging* 2011; **34(1)**: 128-135.

5. Koh D-M ,Collins DJ. Diffusion-Weighted MRI in the Body: Applications and Challenges in Oncology. *Am J Roentgenol* 2007; **188(6)**: 1622-1635.

6. Messiou C, Collins DJ, Morgan VA, Desouza NM. Optimising diffusion weighted MRI for imaging metastatic and myeloma bone disease and assessing reproducibility. Eur Radiol. 2011 Aug; 21(8): 1713-8.

7. Padhani AR, van Ree K, Collins DJ, D'Sa S, Makris A. Assessing the relation between bone marrow signal intensity and apparent diffusion coefficient in diffusion-weighted MRI. AJR Am J Roentgenol. 2013 Jan; 200(1): 163-70.

8. Eschmann SM, Pfannenberg AC, Rieger A, Aschoff P, Müller M, Paulsen F, Anastasiadis A, Claussen CD, Bares R, Schlemmer HP. Comparison of 11C-choline-PET/CT and whole body-MRI for staging of prostate cancer. Nuklearmedizin. 2007; 46(5): 161-8

9. Würslin C, Machann J, Rempp H, Claussen C, Yang B, Schick F. Topography mapping of whole body adipose tissue using A fully automated and standardized procedure. J Magn Reson Imaging. 2010 Feb;31(2):430-9.

10. Jerebko AK, Schmidt GP, Zhou X, Bi J, Anand V, Liu J, Schoenberg S, Schmuecking I, Kiefer B, Krishnan A. Robust parametric modeling approach based on domain knowledge for computer aided detection of vertebrae column metastases in MRI. Inf Process Med Imaging. 2007;20:713-24.

11. Juntu J, Sijbers J, De Backer S, Rajan J, Van Dyck D. Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images. J Magn Reson Imaging.   2010 Mar;31(3):680-9.

12. Pauly O, Glocker B, Criminisi A, Mateus D, Martinez-Möller A, Nekolla S, Navab N. Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences. In

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

Medical Image Computing and Computer Assisted Intervention (MICCAI) 2011.

13. Glocker B, Pauly O, Konukoglu E, Criminisi A. Joint Classification-Regression Forests for Spatially Structured Multi-Object Segmentation. In European Conference on Computer Vision (ECCV) 2012.

14. Wolz R, Chu C, Misawa K, Mori K, Rueckert D. Multi-organ Abdominal CT Segmentation Using Hierarchically Weighted Subject-Specific Atlases, in Medical Image Computing and Computer-Assisted Intervention-MICCAI 2012. 2012, Springer Berlin Heidelberg. p. 10-17.

15. Glocker B, Zikic D, Konukoglu E, Haynor DR, Criminisi A. Vertebrae Localization in Pathological Spine CT via Dense Classification from Sparse Annotations. In Medical Image Computing and Computer Assisted Intervention (MICCAI) 2013.

16. Zikic D, Glocker B, Konukoglu E, Criminisi A, Shotton J, Demiralp C, OThomas OM, Das T, Jena R, Price S. Decision Forests for Tissue-specific Segmentation of High-grade Gliomas in Multi-channel MR. In Medical Image Computing and Computer Assisted Intervention (MICCAI) 2012.

17. Criminisi A, Shotton J, Konukoglu E, Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi- Supervised Learning. 2012, Cambridge: NOW Publishers.

18. Machin D, Campbell MJ, Fayers PM. Sample size tables for clinical studies. Wiley Blackwell Science,2nd ed., 1997

Other relevant literature:

Jiang XX, Yan ZX, Song YY, Zhao WL. A pooled analysis of MRI in the detection of bone marrow infiltration in patients with malignant lymphoma. *Clinical Radiology* 2013; **68(3)**: e143-e153.

Sinkus R, Van Beers BE, Vilgrain V, DeSouza N, Waterton JC. Apparent diffusion coefficient from magnetic resonance imaging as a biomarker in oncology drug development. *European Journal of Cancer* 2012; **48(4)**: 425-431.

Takahara T, Imai Y, Yamashita T, Yasuda S, Nasu S, Van Cauteren M. Diffusion weighted whole body imaging with background body signal suppression (DWIBS): technical improvement using free breathing, STIR and high resolution 3D display. *Radiat Med* 2004; **22(4)**: 275-82.

Kwee TC, Takahara T, Ochiai R, Koh D-M, Ohno Y, Nakanishi K, Niwa T, Chenevert TL,

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019

Luijten PR, Alavi A. Complementary Roles of Whole-Body Diffusion- Weighted MRI and 18F-FDG PET: The State of the Art and Potential Applications. *J Nucl Med* 2010; **51(10)**: 1549-1558.

Houssami N, Costelloe CM. Imaging bone metastases in breast cancer: evidence on comparative test accuracy. Ann Oncol. 2012 Apr;23(4):834-43.

Juntu J, De Schepper AM, Van Dyck P, Van Dyck D, Gielen J, Parizel PM, Sijbers J. Classification of Soft Tissue Tumors by Machine Learning Algorithms, Soft Tissue Tumors, Prof. Fethi Derbel (Ed.), ISBN: 978-953-307-862-5, InTech, Available from: http://www.intechopen.com/books/soft-tissue-tumors/classification-of-soft-tissue-tumors-by-machine-learning- algorithms

Development and evaluation of machine learning methods in whole body MR with diffusion weighted imaging for staging of patients with cancer. Short title: MAchine Learning In whole Body Oncology (MALIBO)

V4.0 21 August 2019