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: 2, **Date:** 22 January 2015

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 Stuart Taylor, Dr Shonit Punwani, Prof Vicky Goh, Dr Dow-Mu Koh, Dr Nishat Bharwani, Dr Sandy Beare

Clinical Queries

Clinical queries should be directed to Prof. Andrea Rockall who will direct the query to the appropriate person

Table of Contents

Keywords and abbreviations

Abbreviations:

MRI (Magnetic Resonance Imaging)

WB DW-MRI (whole body diffusion weighted imaging)

ML: Machine Learning

RT: Reading Time

Keywords:

Normal range

Anatomical whole body imaging

Machine learning

Computer-aided detection

Metastases

Diagnostic test accuracy

Reading time

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 RT 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? 4) What is the best combination of MR sequences for optimal diagnostic accuracy and can this change/be improved by the use of ML techniques? 5) What is the possible change in the number of additional or unnecessary imaging/confirmatory tests used if ML were applied based on the tests that were used for standard staging using WB MRI without ML and at what cost?
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 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. **Secondary objectives:** 1.To compare the RT of WB MR scans; 2. To assess inter-observer variability; 3. To test the diagnostic accuracy of non-experienced readers; 4. To evaluate different combinations of acquired MRI sequences; all the above with and without ML support; 5. To assess the number of possible additional staging tests that would be unnecessary if ML methods were applied with simple cost effectiveness analysis.

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) or physiological

appearances 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 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 applicable tool within the NHS, a method that could assist the radiologist both in diagnostic accuracy and reading speed would be beneficial to deliver improved 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 per-patient 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 unsuspicious regions and to discriminate malignant and 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.

2.2 Clinical studies that this proposal relates to:

The proposed research will use whole body magnetic resonance (WB MRI) data predominantly from the HTA NIHR-funded 'STREAMLINE study'. STREAMLINE is a multi-centre prospective cohort study that evaluates WB MRI in both newly diagnosed non-small cell lung cancer (250 patients) and colorectal cancer (322 patients). The study initially defined WB MRI acquisition, quality assurance and analysis protocols applicable to daily NHS practice. 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; whether early WB-MR could reduce or replace conventional investigations; accuracy of WB MRI and standard diagnostic pathways for local and distant staging based on patient follow-up data; contribution of different MR sequences to the staging accuracy and reporting time; intra- and inter-observer variation in the evaluation of WB-MR datasets by radiologists. The primary outcome measures are the per-patient detection rate for metastatic disease by WB MRI compared to standard imaging diagnostic pathways in non-small cell lung and colon cancer. Secondary outcomes include: influence of WB MRI on time to definitive staging and first major treatment decision, compared to standard diagnostic pathways; comparative total test burden generated by standard imaging pathways and WB-MR; additional tests generated by WB-MR; diagnostic accuracy of WB-MR and standard pathways for metastasis. 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 study started on 1.10.2012 and is planned to end on 1.4.2017. By the end of December 2013, STREAMLINE had recruited 68 patients. The current estimated total recruitment by September 2014 is 160 and by September 2015 is 250. Therefore by September 2016, 250 patients will have completed the study reference standard. The ISRCTN for STREAMLINE-L is ISRCTN50436483 and for STREAMLINE-C is ISRCTN43958015 (1).

Additional cases will be obtained from the CRUK funded MELT study (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). There is infra-structure support for this study through the UCL NIHR Biomedical Research Centre grant. Data from the UCL MASTER study, including cases with lymphoma and prostate cancer will also be used (the myeloma cases from MASTER are unlikely to be used). Planned enrolment, which is nearly complete, 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 one year). 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 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 of WB MRI scans; 2. To assess inter-observer variability; 3. To test the diagnostic accuracy of non-experienced readers; 4. To evaluate different combinations of the acquired MR sequences; all the above with and without ML support; 5. To assess the number of possible additional staging tests that would be unnecessary if ML methods were applied with simple cost effectiveness analysis.

4.0 Study design

This is an observational study 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. 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, MELT and MASTER

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. 150 WB-MR scans will be read by two expert readers over the course of Phase 2, providing a total of 300 radiology reads. Initial radiology reads (approximately the first 100 reads) will focus on identifying ML errors and identification of true positive lesions according to the 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 (the next 100 reads) to develop ML algorithm 'B'. Final Phase 2 reads (the final 100 reads) 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 patient datasets (to allow for sufficient positive cases). 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.

Phase 3 'Validation set': A 2nd set of WB MRI data relating to 217 subjects from the STREAMLINE, MELT and MASTER studies will be read by two expert readers with ML'C' support. The 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 specificity 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. 30 repeat reads (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.

5.0 Eligibility and study population

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

Inclusion criteria

1. Aged 18 to 100 years
2. Written, informed consent
3. Female subjects of reproductive potential should also employ an effective method of birth control. Barrier contraceptives must be used throughout the study in both sexes.

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 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 C and L study inclusion criteria: histopathologically confirmed 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 pregnancy/nursing

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

6.0 MRI protocols

Scanning protocol for Phase 1:

Localiser images will be acquired and then breath-hold anatomical scans (T_1-w with DIXON and T_2-w) will take place in four stations. The patient will be instructed through the intercom to breath in-breath out and hold several times in each station. The table will move between stations. After this the diffusion-weighted scan (with five b-values) will take place, which is free-breathing. The table will again move 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

Previous ML methods will be adapted to WB DW-MRI to allow automatic vertebrae localization, to automatically exclude false positives in unsuspicious 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.

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 order to construct the anatomical atlases, whole body MRI data sets from 50 healthy volunteers will be used. → ML algorithm 'A'.

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'.

8.0 Imaging assessment and patient follow-up

There will be 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 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.

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 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: π= 0.96*0.14+0.86*0.04=0.169
- Odds ratio of the two methods: ψ=0.96*0.14/0.86*0.04=3.9
- Type I error: one side α=0.05 (We believe that WB MRI will be superior with ML support than without ML support)
- Power: 1-β=90%

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 80% based on the preliminary data of STREAMLINE study, the total sample size for Phase 3 is 177 patients from STREAMLINE. Therefore, 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%.

Among the 177 patients from STREAMLINE study, there are 36 metastases. The MELT and MASTER studies can further provide 40 patients for Phase 3 (most of whom will have multiple sites of disease), so the total number of metastases would be 76. Assuming the sensitivity of WB-MR with ML support is no less than that of WB-MR alone (88%) while the specificity can improve 10% from 86%, with a sample size of 76 metastases, the 95% confidence interval for the sensitivity of WB-MR with ML support is expected to be 81% to 95%. Thus the final sample for Phase 3 is 217 patients.

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.The sensitivity 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. Sensitivity is 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 without ML‘C’ support; 2. Reading time of WB-MRI with or without ML ‘C’ support 3. Inter-observer variability: measured by the Kappa Coefficient between the experienced radiologists; 4. Diagnostic accuracy of inexperienced readers WB MRI with or without ML ‘C’ support: measured by per patient specificity and sensitivity against the reference standard established in the main study; 5. Diagnostic accuracy of limited WB MRI sequences (T1 and DW-MR) will be measured by sensitivity and specificity against reference standard with or without ML support; 6. Cost-effectiveness: measured by the cost of applying ML support against the time and resource (number of possible additional staging tests) saved after applying the ML 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 the two experienced (and inexperienced radiologists in a sub-dataset of 30 patients) 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 among a sub-dataset of 30 patients (see 4. Study design).

Diagnostic accuracies of limited WB-MR sequences: Diagnostic accuracies of limited WB-MR sequences will be summarised by sensitivity and specificity with 95% confidence interval and will be compared with the diagnostic accuracy of full multi-sequence WB-MR.

Simple Cost-effectiveness: Cost-effectiveness will be summarised by the per patient cost of applying ML support and the per-patient time and resource (number of possible additional staging tests) saved after applying the ML support taking consideration of different prevalence of metastasis.

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 and MASTER studies. Patients with missing data or loss to follow-up will not be 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 study). 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 phase 1 is in place. Ethical approval for phases 2 and 3 will now be requested. We understand it is highly likely that the study will be approved with proportionate review as we will be reviewing data that has already been obtained, no patient identifiable data will be available to the researchers and no patient intervention or change in patient management will take place as a result of the study. 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 sponsor will be Imperial College. The Trial MR physicist will manage the study day to day, with clinical trials unit (CTU) oversight from a part-time clinical trials manager.

The physicist will coordinate R & D approval for Phase 2 and 3 together with the CTU, and they will ensure trial and investigator documentation is in order and version controlled. The MR physicist will (i) work with D Rueckert and B Glocker to ensure appropriate segmentation of all MR datasets for ML analysis (ii) collaborate with Biotronics 3D team to ensure appropriate anonymised dataset transfer between the ML computer interface and the Biotronics 3D radiology reading interface (iii) will ensure that the study documentation and trial master file, draft reports to funders and ethics committee will be up to date and completed. The study investigators will regularly supervise the study progress to ensure the timetable is met, that progress is satisfactory. The UCL CTC will work with the study applicants to adapt the database to the needs of the proposed mechanistic study and ensure integrity of the data. The Trial Management Group (TMG) will oversee the project with one meeting and three telephone conferences per year. This group will include the CI and PIs, CTU statistician and the PPI representative. They will identify and address any concerns and decide on any necessary actions to keep the study on target. A trial steering committee (TSC) will meet by telephone conference twice yearly to provide independent oversight of the study and make recommendations as necessary. We have provided letters of support from the CI's, sponsors and funders of STREAMLINE, MELT and MASTER studies. There are no proposed changes to the main study should the mechanistic study be funded.

11.0 Study management

The trial MRI physicist will manage the study day to day, with clinical trials unit oversight with the part-time clinical trials co-ordinator from ICTU. The CI and physicist will make the ethical application for Phases 2 and 3 prior to the planned commencement of the study, planning to have ethics approval in place at start. The physicist will coordinate the R&D approval for Phase 2 and 3 and together with the TC, ensure trial and investigator documentation are in place. The MRI physicist will work with D Rueckert and B Glocker to ensure appropriate segmentation of all MRI datasets for ML analysis. MR physicist will collaborate with the management team of the secure server (Biotronics 3D) to ensure appropriate anonymised dataset transfer to the ML computer interface. The MR physicist together with TC and CTU support will ensure the study documentation and trial master file, draft reports to funders and ethics committee will be up to date and completed. The study investigators will regularly supervise the study progress to ensure the timetable is

met and that progress is satisfactory. The TMG will oversee the project. This group will include the CI and PIs, the CTU statistician and the PPI representative. The group will meet quarterly to up-date on progress and identify any concerns that need to be addressed. They will decide on any necessary actions to keep the study on target. A trial steering committee (TSC) will provide independent oversight of the study and make recommendations as necessary.

12.0 Project timetable and milestones

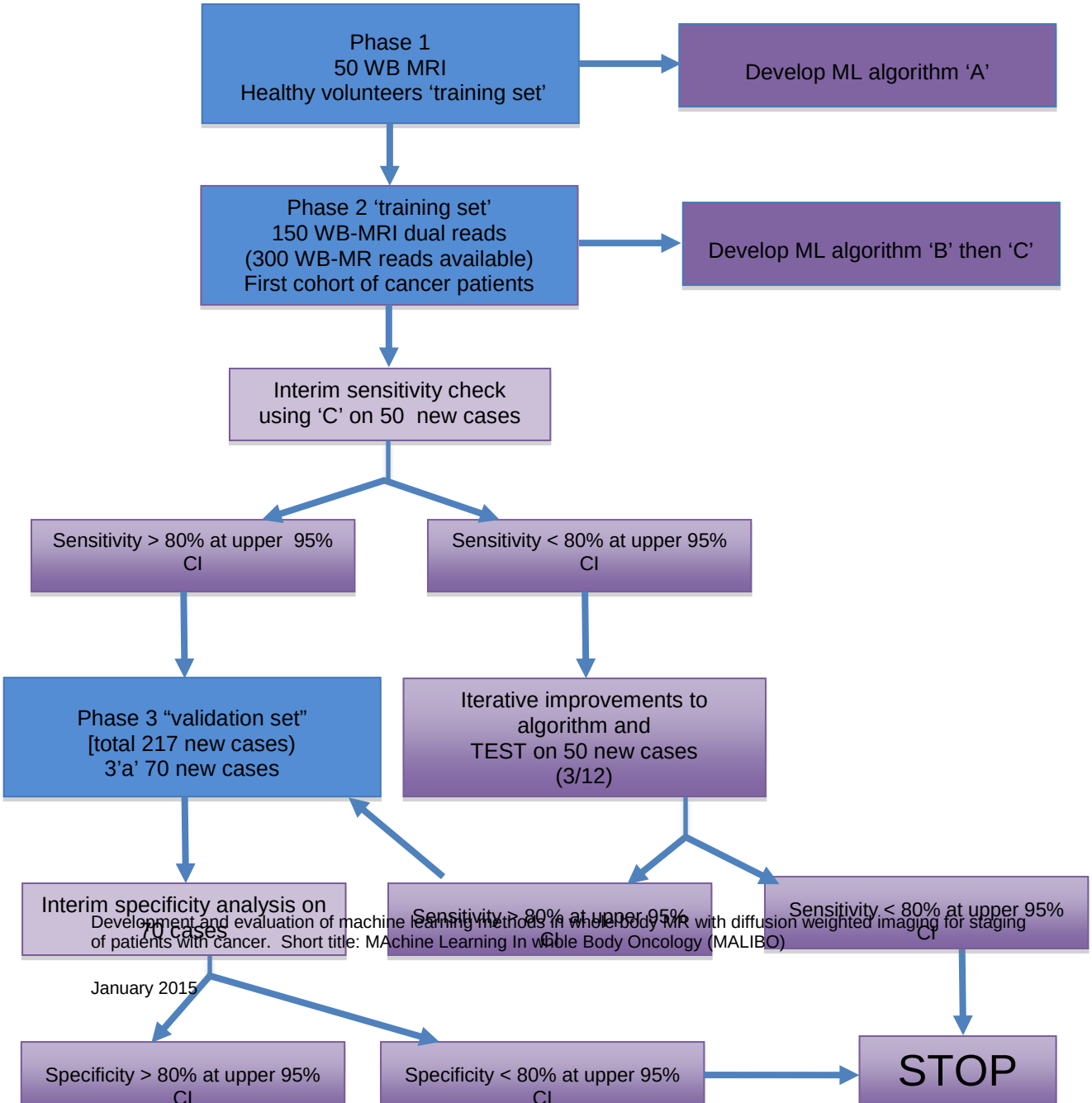
We estimate the study start date to be Feb 1 2015.

Phase 1, Months 0-9: Study set up and development of ML algorithm 'A'. Month 1: transfer of 50 anonymised healthy volunteer datasets to ML computer platform (Milestone 1). **Months 1-3:** Set-up co-ordination with secure central server in preparation for phase 2; R&D approval and ethics complete for Phase 2 and 3 (Milestone 2). **Months 2–6:** segmentation of all healthy volunteer datasets (Milestone 3); **Months 3-9:** development of ML algorithm 'A' (Milestone 4).

Phase 2, Months 10-29: First cohort (training set) of radiology reads and refinement of ML algorithm. 150 cases from STREAMLINE, MELT and MASTER studies will be read by two experienced WB-MR readers to delineate the known lesions agreed by reference standard, thus a total of 300 reads available for analysis (each of the six readers will read 50 scans). **Months 10-12:** First 100 reads for identification of tumour sites using ML'A' with reader identification of errors. Milestone 5 (complete 100 reads). Months 11-13: refine ML'A' to develop ML 'A+'. As reads become available, they will start to feed in to refinement of ML algorithm. The final reference standard will be used to confirm sites of disease for these patients. **Months 12-15:** write up phase 1. Milestone 7 submit phase 1 study. **Months 14-16:** Second set of 100 reads. Milestone 8 (complete 2/3 phase 2 reads). **Months 15-18:** Refinement of ML'A+' to develop ML algorithm 'B'. Milestone 9, complete ML 'B'. Months 19- 21: complete radiology reads. Milestone 10 complete reads phase 2. **Months 19 – 27:** refine ML 'B' to develop ML'C'. Milestone 11, complete ML 'C'. **Months 28-29:** Phase 2 interim analysis 1: detection rate of lesions in 40 new cases.

Phase 3, Months 30-42: Second cohort (validation set) of radiology reads using ML algorithm 'C' and final report. 217 subsequent cases from STREAMLINE, MELT and MASTER studies will be read by two experienced WB-MR readers (total reads 435). **Months 30-36:** Initial 50-70 cases read and interim analysis (allowing 2 months). Complete reading half of phase 3 scans. Data entry. Milestone 12, 50% reads completed. **Months 36-39:** Complete radiology reads. Data entry. Milestone 13, completed radiology reads. **Months 33-39:** Sub-study evaluating RT, diagnostic accuracy and inter-observer variation using ML algorithm 'C' in sub-sets of 30 cases with inexperienced readers. Milestone 14 complete sub-study reads. **Months 36-39:** ensure all final reference standards are available STREAMLINE, MELT, MASTER studies (Milestone 15, combining read results with completed reference standards). **Months 39-42:** Statistical analysis for primary and secondary outcome measures. Manuscript preparation and final report submission (Milestone 16).

12.0 Study plan flow diagram



References

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 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, 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