

BMJ Open Development of machine learning support for reading whole body diffusion-weighted MRI (WB-MRI) in myeloma for the detection and quantification of the extent of disease before and after treatment (MALIMAR): protocol for a cross-sectional diagnostic test accuracy study

Laura Satchwell ¹, Linda Wedlake,¹ Emily Greenlay,¹ Xingfeng Li,² Christina Messiou,^{1,3} Ben Glocker,⁴ Tara Barwick,^{2,5} Theodore Barfoot,⁶ Simon Doran,³ Martin O Leach,³ Dow Mu Koh,^{1,3} Martin Kaiser,^{1,3} Stefan Winzeck,⁴ Talha Qaiser,⁴ Eric Aboagye,² Andrea Rockall^{2,5}

To cite: Satchwell L, Wedlake L, Greenlay E, *et al*. Development of machine learning support for reading whole body diffusion-weighted MRI (WB-MRI) in myeloma for the detection and quantification of the extent of disease before and after treatment (MALIMAR): protocol for a cross-sectional diagnostic test accuracy study. *BMJ Open* 2022;**12**:e067140. doi:10.1136/bmjopen-2022-067140

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-067140>).

Received 03 August 2022
Accepted 25 August 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Laura Satchwell;
laura.satchwell@rmh.nhs.uk

ABSTRACT:

Introduction Whole-body MRI (WB-MRI) is recommended by the National Institute of Clinical Excellence as the first-line imaging tool for diagnosis of multiple myeloma. Reporting WB-MRI scans requires expertise to interpret and can be challenging for radiologists who need to meet rapid turn-around requirements. Automated computational tools based on machine learning (ML) could assist the radiologist in terms of sensitivity and reading speed and would facilitate improved accuracy, productivity and cost-effectiveness. The MALIMAR study aims to develop and validate a ML algorithm to increase the diagnostic accuracy and reading speed of radiological interpretation of WB-MRI compared with standard methods.

Methods and analysis This phase II/III imaging trial will perform retrospective analysis of previously obtained clinical radiology MRI scans and scans from healthy volunteers obtained prospectively to implement training and validation of an ML algorithm. The study will comprise three project phases using approximately 633 scans to (1) train the ML algorithm to identify active disease, (2) clinically validate the ML algorithm and (3) determine change in disease status following treatment via a quantification of burden of disease in patients with myeloma. Phase 1 will primarily train the ML algorithm to detect active myeloma against an expert assessment ('reference standard'). Phase 2 will use the ML output in the setting of radiology reader study to assess the difference in sensitivity when using ML-assisted reading or human-alone reading. Phase 3 will assess the agreement between experienced readers (with and without ML) and the reference standard in scoring both overall burden of disease before and after treatment, and response.

Ethics and dissemination MALIMAR has ethical approval from South Central—Oxford C Research Ethics Committee (REC Reference: 17/SC/0630). IRAS Project ID: 233501.

STRENGTHS AND LIMITATIONS OF THIS STUDY:

- ⇒ The MALIMAR study has the potential to acquire and characterise what is possibly the largest set of myeloma WB-MRI scans in the UK.
- ⇒ The cross-sectional diagnostic accuracy design allows for retrospective analysis of previously obtained clinical radiology scans for training and validation of an ML algorithm.
- ⇒ This study will provide ML outputs that can be tested across the National Health Service in live real-time clinical settings.
- ⇒ As data will be acquired over a long period of time, scan quality could vary.
- ⇒ Replicating clinical reporting in a retrospective study setting can be difficult to achieve, particularly for analysis of scan reading time.

CPMS Portfolio adoption (CPMS ID: 36766). Participants gave informed consent to participate in the study before taking part. MALIMAR is funded by National Institute for Healthcare Research Efficacy and Mechanism Evaluation funding (NIHR EME Project ID: 16/68/34). Findings will be made available through peer-reviewed publications and conference dissemination.

Trial registration number NCT03574454.

INTRODUCTION

There is strong evidence in the existing literature for the use of whole-body MRI (WB-MRI) in the management of patients with multiple myeloma. In 2016, the National Institute of Clinical Excellence (NICE) made the recommendation of using WB-MRI as the first-line

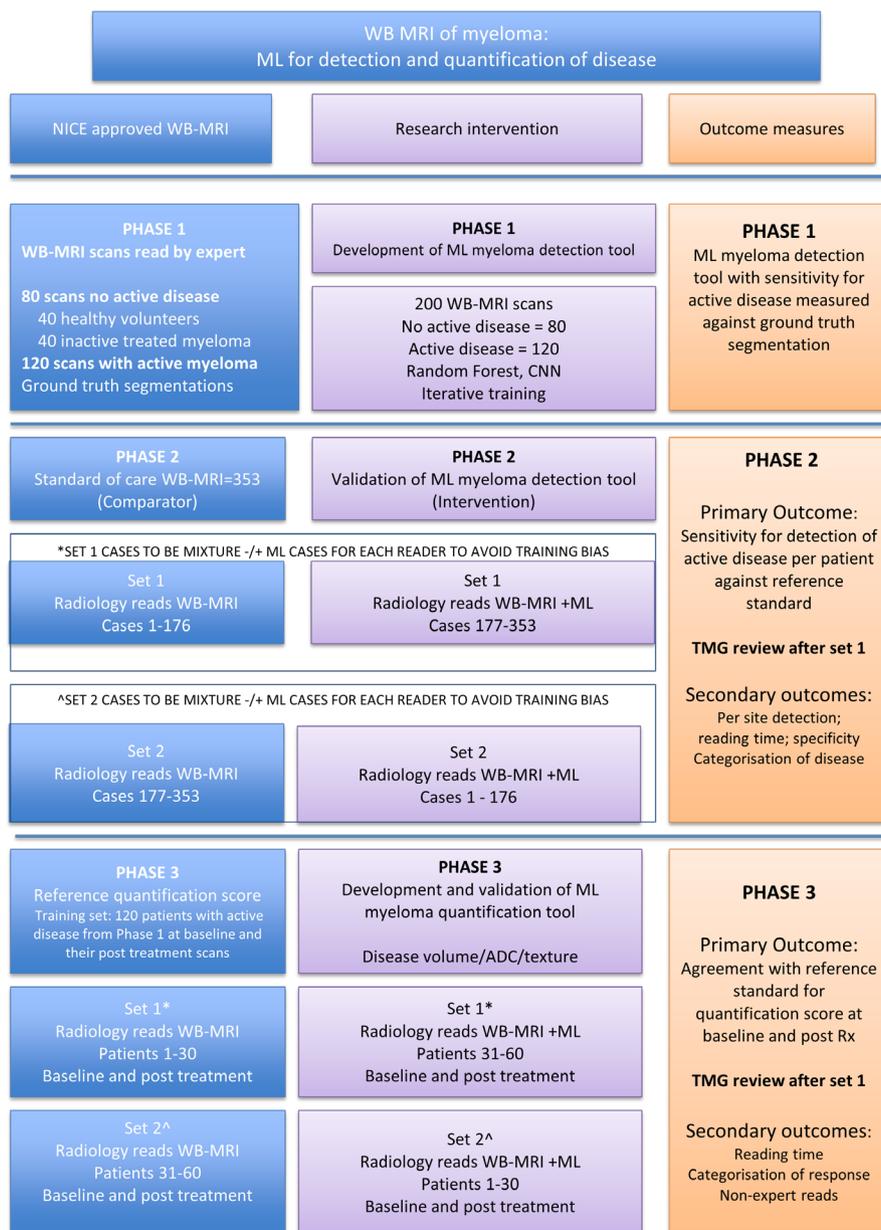


Figure 1 MALIMAR study flow diagram. ADC, apparent diffusion coefficient; CNN, convolutional neural network; ML, machine learning; NICE, National Institute of Clinical Excellence; TMG, trial management group; WB-MRI, whole-body MRI.

imaging tool for diagnosis, based on the literature.¹ A consensus from the International Myeloma Working Group agreed that identification of focal lesions more than 5 mm on MRI should now be used as an indication to treat.^{2,3} Evidence suggests that diffusion-weighted (DW) WB-MRI (WB-DW-MRI) is the most sensitive magnetic resonance technique for detecting marrow disease⁴⁻⁸ and superior to fluorodeoxyglucose positron emission tomography/CT for the detection of small sites of disease and diffuse infiltration.^{9,10} Therefore, WB-MRI is increasingly being adopted at centres worldwide for patients with myeloma. Treatment of high-risk patients is known to improve overall survival,¹¹ therefore improved diagnostic accuracy is likely to translate into improved patient selection for treatment and prolonged survival.

Despite the acknowledged benefits of WB-MRI for patients with myeloma, with publication of the NICE guidance, one of the major concerns is how these complex scans can be reported by a radiology workforce in crisis. Specificity of disease detection in the marrow is improved by viewing source DW images alongside quantitative apparent diffusion coefficient (ADC) maps. This allows differentiation of active sites of disease with restricted diffusion from treated sites of disease and vertebral haemangiomas, which conversely return a very high ADC.¹² Dixon images are also integral to image interpretation and morphological imaging is also necessary to identify mechanical complications of myeloma bone disease. Therefore, diagnostic accuracy is dependent on viewing multiple imaging sequences⁷ and typically over

Table 1 Comparison of MALIMAR anatomical regions between ground truth CRFs and reader CRFs

Anatomical regions	
Ground truth CRFs (phases 1 and 2)	Reader CRFs (phase 2)
Skull	Skull
Scapula right	Ribs/clavicles/sternum/scapulae
Scapula left	Ribs/clavicles/sternum/scapulae
Clavicle right	Ribs/clavicles/sternum/scapulae
Clavicle left	Ribs/clavicles/sternum/scapulae
Sternum	Ribs/clavicles/sternum/scapulae
Spine upper	Cervical spine
Spine middle	Dorsal spine
Spine lower	Lumbar spine
Ribs right	Ribs/clavicles/sternum/scapulae
Ribs left	Ribs/clavicles/sternum/scapulae
Sacrum	Pelvis
Femur right	Long bones
Femur left	Long bones
Humerus right	Long bones
Humerus left	Long bones
CRFs, case report forms.	

1200 image slices per WB-MRI scan in order to achieve whole body coverage. Consequently, reading time for the scans may be significant. At least 9% of UK radiology posts are unfilled,¹³ and in 2015, clinical radiology was placed on the national shortage occupation list. The time-consuming process of reporting WB-MRI scans is a concern for radiologists who need to provide rapid turn around with a high productivity to support the National Health Service (NHS). Automated computational tools based on machine learning (ML) could support reporting of these large data sets and facilitate translation of this valuable imaging technique into the NHS, not only in detecting active disease but also in identifying response to treatment. Ideally, an ML algorithm would automatically detect and highlight suspicious regions and could reduce reading time. An accurate and automatic detection of pathology may also increase diagnostic accuracy.

The possibility of using computer-assisted ML techniques has been considered in aiding interpretation of complex imaging data sets.^{14–16} Current work in the EME NIHR (Efficacy and Mechanism Evaluation National Institute of Health Research) funded MALIBO study^{17 18} (13/122/01) has demonstrated fully automatic multiorgan segmentation using WB-MRI in healthy volunteers (HV) and ML detection of primary colorectal cancer and metastatic lesions.

Aim

The aim of the MALIMAR study is to develop and validate an ML algorithm to improve the sensitivity of

radiologists to detect the presence and extent of active myeloma before and after treatment, with high reproducibility and reduced reading time (WB-MRI with ML, the intervention) when compared with the standard of care radiology read (WB-MRI without ML support, the comparator).

METHODS AND ANALYSIS

Study design

The study is based on a cross-sectional diagnostic test accuracy design and will comprise three distinct project phases as summarised in [figure 1](#).

- ▶ In phase 1, the ML algorithm will be trained using both HV and myeloma patient scans to recognise active myeloma deposits as distinct from cases with no active disease, classifying disease as ‘focal’, ‘diffuse’ or ‘inactive’.
- ▶ In phase 2, the ML algorithm will be validated using a second unseen data set against a reference standard (ie, ground truth) to assess how accurately radiologists classify disease using scans with the ML algorithm and compared with readings without ML. Diagnostic accuracy on a per patient and per region (using 16 predefined anatomical sites—[table 1](#)) basis and reading time will be measured.
- ▶ In phase 3, further development of the ML algorithm to quantify disease burden will be undertaken using data sets from phase 1 and 2. This quantification output will be tested in the phase 3 reader study in which readers will record disease burden and response between paired baseline (new diagnosis or relapse prior to initiation of treatment) and single post-treatment WB-MRI scans, with or without ML support, and tested against the reference standard.

Participants and recruiting centres

The study will be run at The Royal Marsden NHS Foundation Trust across two Royal Marsden Hospital (RMH) sites; Chelsea and Sutton and Imperial College Healthcare Trust (ICHT). Patient and HV scans will make up the study population, and disease classification will be at both the scan and anatomical site level.

The scan population will comprise of; HV WB-MRI scans acquired from participants prospectively recruited from the sponsor site only (RMH), with the option of the Imperial Site providing previously acquired HV scans; WB-MRI scans acquired as part of clinical care from patients being managed at RMH and ICHT and WB-MRI scans previously acquired for a prospective research study in WB-MRI (iTMM study).^{9 19} All scans acquired for the study will be done, so using clinical standard of care trust protocols.

The inclusion/exclusion criteria for the HV and patient scans are detailed in [table 2](#) and the planned number of scans for each study phase is detailed in [table 3](#).

Table 2 Inclusion and exclusion criteria

	Inclusion criteria	Exclusion criteria
Healthy volunteers	Written informed consent No contra-indication to MRI 40 years or above in age (attempts will be made to include similar age range as myeloma patients) No known significant illness No known metallic implant	Significant artefact on scan Corrupted scan data
Patients in phases 1 and 2	Patient with confirmed myeloma with WB-MRI scan previously performed as part of clinical care. Sufficient imaging and clinical data for the expert reference panel to categorise the WB-MRI scan as: 1. Previously treated inactive disease with no evidence of active disease based on expert reference panel 2. Active disease—focal 3. Active disease—diffuse 4. Active disease—extra-medullary 5. New active myeloma, no previous treatment Patients may be included if the pattern of disease is a combination of focal, diffuse and/or extra-medullary.	Corrupted WB-MRI scan data. Insufficient clinical data to allow the expert reference panel to categorise the scan.
Patients in phase 3	Training set: phase 1 active disease cases and their post-treatment scans from phase 2. Validation set: from iTIMM study. Written informed consent for iTIMM study All patients over the age of 18 with multiple myeloma planned for autograft.	Corrupted scan data. MRI incompatible metal implants Claustrophobia Diagnosis of other malignancy within 5 years

iTIMM, Image-guided Theranostics in Multiple Myeloma; WB-MRI, Whole-Body Magnetic Resonance Imaging.

Intervention and reference standard

Intervention (including comparator)

The comparator in this study is defined as WB-MRI scans read by experienced radiologists, as per standard care (WB-MRI, the COMPARATOR). The intervention will use these standard methods with the addition of ML (WB-MRI+ML, the INTERVENTION). The ML algorithm will be developed during phase 1 of the study following data curation and scan allocation to phases 1 and 2. DW imaging, ADC map and T1-weighted sequences (Dixon fat and water scans) will be used, reflecting the radiological reading tools used by expert readers.

Radiologists or readers are defined as experienced based on their previous clinical radiology reading skills and responsibilities and their length of service in this role.

Experienced readers will be required to have completed at least 100 WB-MRI clinical scan reports.

Reference standard

There is no available histological reference standard for every site of bone marrow disease, as trephine biopsy is usually restricted to a single site. The proposed reference standard, thus, comprises the interpretation of an expert panel; a radiologist and a haematologist who are experts in myeloma. They will have access to (1) WB-MR images, (2) bone marrow histopathology reports (with quantitation), (3) serum paraproteins, (4) serum-free light chain (sFLC), in order to categorise per scan:

- ▶ Presence or absence of active disease.
- ▶ The detailed disease distribution by anatomical site.

Table 3 Number of healthy volunteer (HV) and multiple myeloma (MM) scans in each category for each study phase

	HV*	MM inactive	MM active focal	MM active diffuse	MM new diagnosis	Total
Phase 1†	40	40	60	40	20	200
Phase 2	50	100	105	70	28	353
Phase 3 training‡	0	(80 post-treatment)	60	40	20	200
Phase 3 validation	0	60 patients in iTIMM study scanned at baseline and post-treatment				120

*A total of 50 HV will be used, 40 in phase 1, which will be used again in phase 2, with the addition of 10 more HV.

†The number of scans in phase 1 may increase by 140–180 scans (100 subjects) if there is evidence of over-fitting in the development of the algorithm.

‡Scans used in phase 3 training are scans that have been previously used in phases 1 and 2.

HV, Healthy Volunteer; iTIMM, Image-guided Theranostics in Multiple Myeloma; MM, Multiple Myeloma.

- ▶ Quantitation of the burden of disease (using a validated MRI score^{20 21} and sFLC) including category of response to treatment .

Scan and site-level data from these scans will be captured on case report forms (CRFs) for all cases in phases 1 and 2 and used as ‘ground truth’ in the classification of study output. Reference standard for phase 3 will be obtained from the source (iTMM study).¹⁹

Objectives

Primary research objectives

Phase 1: to develop a myeloma-specific ML algorithm to detect the presence of active disease on WB-MRI+ML (with machine learning ‘+ML’) with sufficient sensitivity.

Phase 2: to validate WB-MRI+ML against the comparator WB-MRI for sensitivity on a per-patient and per-site basis.

Phase 3: to develop and validate an ML algorithm to automatically quantify the burden of active disease, before and after treatment.

Secondary research objectives (phases 2 and 3 only)

For each of the following, our objective is to compare WB-MRI with and without ML support to the reference standard for:

1. Reading time.
2. Specificity.
3. Sensitivity of non-experienced readers.
4. Agreement of categorising disease as focal, diffuse and/or extramedullary.
5. Agreement of categorising patients as responder or non-responder.

Procedure

Scan acquisition—HV

HV will be recruited to obtain data from normal bone marrow within the age range typical of myeloma. Up to 50 HVs aged 40 years or above will be recruited using approved advertisements at the Sponsor site and consented with the help of clinical research network (CRN) resources (see online supplemental file 1a for consent form). The HV information sheet (online supplemental file 1b) will clearly explain the MRI scanning procedure and the actions that will be taken in the event of incidental (ie, unexpected) findings. Contact details will be supplied on the HV information sheet to enable volunteers to respond to the invitation or ask any questions. A total of 22 HV scans previously acquired are also available for use from ICHT if needed.

Participating HVs will undergo a single whole body MRI scan at RMH according to the trial-specific scanning protocol. HV scans will be acquired in the following sequences (T1, fat/water, Dixon, ADC, etc) to mirror the clinical setting and on Siemens, Avanto and Aero (wide bore) MRI scanners. Subjects with a larger body mass index will be scanned on the Siemens Aero, which has a larger bore diameter to optimise comfort.

Scan acquisition—patients with myeloma

Previously acquired patient scans will be identified by the investigators within the Sponsor’s myeloma clinical service (between 2011 and 2020), supplemented by scans from ICHT, until the required sample size is reached. Scans will normally include the following sequences; T1, fat/water, Dixon, ADC, etc, and on the following MRI machines; Siemens, Avanto and Aero MRI scanners (online supplemental file 2 for sequence details).

Scan classification and allocation to study phase

Patient scans will be categorised by the expert reference panel as showing inactive disease, active focal, active diffuse (focal or diffuse) and new disease. HV scans will be classified as normal (ie, non-diseased). Scans will be allocated to Phase one or two as per table 3. To minimise bias or ‘over-learning’, no more than five scans from the same patient will be allocated to Phase 1. Phase two scans will not include any patient scans that have been used in Phase one and thus comprise only those previously unseen by the ML algorithm. A subset of scans from phase 1 and 2 will be used to further train the algorithm at the start of Phase 3. Phase three validation scans have previously been acquired for the iTMM trial (NCT02403102) and include a unique series of paired scans, previously unseen by the ML algorithm.

Scan curation (quality control) and anatomical segmentation

Eligible scans will be curated immediately prior to transfer to an online platform for secure storage (ICR XNAT). This will ensure that the ML algorithm is able to interpret all scans consistently. Curation scripts will be written in python and ensure that scans exhibit consistent characteristics such as: correct sequential display of images, no missing slices, noting presence of unusual artefacts that might interrupt ML reads and other factors which might compromise interpretation. Further details on the data curation will be published elsewhere.

Phase 1 scans will then be manually segmented into 16 bone regions (table 1) using a boundary box approach. These scans will be used to teach the ML algorithm to recognise active myeloma disease (focal or diffuse) and precision metrics will be evaluated in order to achieve the optimal algorithm. Initially, scans will be classified by the ML algorithm at scan level (ie, patient level) only.

Testing of ML algorithm—radiology reading process

The ML algorithm will be tested by both experienced and inexperienced radiology readers.

Phase 2 scans will be subjected to the ML algorithm, which will provide an ML overlay on all scans, indicating areas of disease by means of a heat map. For each scan, a ‘standard’ and ‘ML’ version will be available. The trial statistician will randomly allocate reads to each of the (approximately 15–20) readers, using trial-specific algorithms written using Stata software (StataCorp, Texas). The reads will be performed in two batches to incorporate a wash-out period. Each batch will have 50% of cases with

ML support and 50% without, to avoid reader training bias. The reading process will be described in a reader manual and all readers will receive appropriate training in viewing scans using the Biotronics 3D web-based platform and completing a Read CRF available via Microsoft Forms (see online supplemental file 3a). In the case of 'inexperienced' readers, training will comprise a review of the CRFs and the viewing software with a basic training on reporting lexicon. A scribe will be provided to assist readers during the reading process and input data to the CRF in each batch of reads. Following a 4-week wash out period, readers will be presented with the second batch of reads with the opposite reading paradigm with regards to the ML support. The same cases will be allocated to the same readers. A subset of approximately 50 scans will be read a second time by a different reader as an inter-rater check.

In phase 3, scans from the iTIMM study, comprising paired baseline and follow-up post-treatment scans, will be used to test whether the ML algorithm is capable of distinguishing change in disease status (ie, disease burden) between the two time points. Reads will again be randomly allocated to the readers by the trial statistician. Readers will follow similar procedures to that outlined above with one set of paired scans having the ML overlay and the other with no ML overlay (for CRF, see online supplemental file 3b). A 4-week wash out period will again apply between the two batches of reads. A subset of approximately 20 scans will be read a second time by a different reader as an inter-rater check.

Data collection

Reader responses will be captured using MS Forms with responses being transferred directly to an excel spreadsheet. Examples of the CRFs to be used in both ML validation phases are given as online supplemental file 3 a,b. All readers will be provided with a manual describing CRF completion (including a lexicon of disease definitions) and use of the software viewing tools and overlay of the ML output heatmap and opportunity for live training using the online platform.

Outcome measures

Phase 1—ML algorithm training phase

Primary: sensitivity for the detection of active myeloma on WB-MRI+ML detection tool against the reference standard.

Secondary: (1) specificity; (2) F1 score (a single measure of precision and recall).

Phase 2—ML algorithm clinical testing phase (presence/absence of active myeloma)

Primary: difference in sensitivity of WB-MRI-/+ML detection tool to diagnose the presence of active myeloma on a per-patient basis, by experienced readers, assessed against the reference standard.

Secondary: for comparison of WB-MRI-/+ML: (1) per-site sensitivity to diagnose active disease, (2) reading time,

(3) specificity, (4) agreement with reference standard to categorise disease as focal, diffuse and/or extramedullary, (5) Sensitivity of non-experienced readers for presence of active disease.

Phase 3—ML algorithm for quantification of disease burden with clinical testing

Primary : agreement between experienced readers and the reference standard in scoring overall burden of disease before and after treatment for response categorisation -/+ ML quantification tool.

Secondary: for comparison of WB-MRI -/+ML: (1) reading time, (2) agreement of categorisation of patients as responder or non-responder with the reference standard, (3) agreement of non-experienced readers for burden of disease and categorisation of response, (4) estimated difference in cost for radiology reading time for WB-MRI -/+ML.

Proposed tertiary: verification of the team's previously published work regarding reverse classification accuracy: predicting segmentation performance in the absence of a reference standard.²²

Sample size

Phase 1

We will train the ML algorithm on a set of scans without and with active disease that will reflect the categories of disease that may be encountered in clinical practice. The number of cases used for training are arbitrarily chosen reflecting the knowledge that a large number of training data sets will improve training accuracy, counterbalanced with the resources needed to curate and annotate a large number of data sets.

Phase 2

The study is powered on the primary outcome of sensitivity.

In a meta-analysis, Wu *et al* have reported a pooled sensitivity of 88% and a pooled specificity of 86% (0.86 for WB-MRI with DW-MRI).⁸ We anticipate that the addition of ML could increase this by at least 7.5%, from 88% to 95.5%. There is no background data to indicate the expected proportion of discordant pairs, so we have estimated this as $(1-0.955) \times 0.88 + 0.955 \times (1-0.88)$, which is equal to 0.154. To achieve 80% power using a two-sided alpha of 0.05 would require a total of 203 patients positive for myeloma using the gold standard.

If it is assumed that the specificity will be unchanged using ML, a total number of cases with no active disease of 150 (50 HV, 100 inactive treated myeloma) will give 80% power to show that the difference is above a non-inferiority limit of 10%.

Phase 3 training

Approximately 200 cases that have at least two time points will be taken from phases 1 and 2, with active disease present at least at one time point, and used for training and validation for burden of disease; this will ensure efficient use of all data and segmentations.

Phase 3 clinical testing

This sample size is fixed at 60 patients, the full sample size of the iTIMM study, each of whom has a baseline and one post-treatment scan.

Statistical analysis

Phase 1 analysis

The ability to correctly localise and detect active disease will be evaluated by calculating sensitivity, specificity and the F1 score (a single measure of precision (positive predictive value) and recall (sensitivity)) for multiple algorithms and compared against the reference standard. Following Trial Steering Committee (TSC) approval, the optimal algorithm will move forward to phase 2.

Phase 2 analysis

In phase 2, the percentage of patients with active disease on WB-MRI+/-ML support who have positive reference standard will be compared using McNemar's test with a two-sided alpha of 0.05. Per-patient and per-site sensitivity and specificity with and without ML support will be reported with 95% CIs. Reading time will be compared using Wilcoxon's test for paired data and described using summary statistics.

The same analysis of sensitivity, specificity and reading time will be repeated for inexperienced readers.

Agreement between experienced and inexperienced readers will be measured in a subset of cases with a Kappa coefficient and overall proportion of concordant cases.

All other endpoints will be summarised using descriptive statistics.

Although the study is powered to detect superiority of the primary endpoint, if sensitivity is shown to be non-inferior using ML and reading time is both clinically and statistically significantly lower using ML, this would be considered as an indication to proceed. Non-inferiority in this context will be defined as having any possible reduction in sensitivity with ML significantly higher than a lower limit of -10% (using Tangos' test with one-sided alpha of 0.05).

Phase 3 analysis

In phase 3, the difference between the experienced readers' disease score to the reference standard disease score will be recorded and compared +/-ML support using Wilcoxon's test. Differences from scores given by experienced readers and the reference standard will be described using Bland-Altman plots for scores \pm ML support.

All other endpoints will be summarised using descriptive statistics.

A simple cost-effectiveness analysis may be performed depending on study findings, such as the reading time.

Procedure(s) to account for missing or spurious data

If a scan is incomplete or the file is corrupted and not evaluable, it will be excluded from the data set. If a set of radiology reads is incomplete, a new trained reader will be identified to do the full allocation of reads.

Timing and responsibility for analyses

Analyses will take place at both the end of phase 2 and then again at the end of phase 3, when all readings have been completed.

Patient and public involvement

A patient and public involvement (PPI) representative was appointed from an established group at Myeloma UK. The individual gave in-depth feedback on the study, particularly on the relevance to patient care and the use of retrospective patient data and HV scans. Myeloma UK is fully supportive of the project and is willing to assist with dissemination of important findings to the Myeloma UK community.

Safety

As this study is recruiting HV only, an a priori agreement has been reached with the sponsor that safety reporting is not required. Sponsor procedures in respect of incidental (ie, unexpected) findings in HV will be adhered to and results were captured within the Trust's Clinical Record.

Monitoring against Source Data will not be required, which is in line with the Sponsor's policy on non-Clinical Trial of Investigational Medicinal Product trials.

Trial funding, organisation and administration

The study has been awarded funding by Medical Research Council NIHR EME Awards Body (NIHR EME Project ID: 16/68/34). In addition, the department of radiology has agreed to fund the cost of HV WB-MRI scans. The cost of recruitment and consenting of HVs will be requested through the NHS CRN. RMH is the study sponsor responsible for initiating and managing the study and the coordinating centre, including sign-off of the study protocol.

A trial management group (TMG) meeting will be held regularly to ensure satisfactory progress of the study. A TSC will provide independent oversight for the study, review the development of the ML algorithm and advise the TMG where problems may arise. The TSC will include a patient advocate.

Ethics and dissemination

Ethical approval for MALIMAR was granted on 21/11/2017 (REC) and 21/12/2017 (Health Research Authority) Here, we report V.3.0 of the protocol. All participating sites gained local approval prior to study participation.

Any protocol modifications will be submitted for approval to the REC, reflected in the online registration and disseminated by e-mail to site principal investigators and trial coordinators. The statistician will have access to the final linked trial data set. There are no plans to provide public access to the full protocol, participant-level data or statistical code. The researchers aim to publish results in a peer-reviewed journal and share via social media and conferences. Authorship will be determined according to academic standards.



DISCUSSION

This study aims to develop and validate an ML algorithm to augment the performance and efficiency of the radiology reading process using WB-MRI. The results will show the impact of using the ML tool and outcomes of the study will have implications for the application of ML with WB-MRI in patients with patients across the NHS. It is anticipated that feasibility analysis will follow the successful completion of this study to pilot the implementation of the ML tool in a real-time prospective study prior to future clinical setting.

To avoid bias, we ensure: (1) comparator and intervention tests are read by readers that are fully blinded to the reference standard, (2) a mixture of cases with and without disease, (3) the reads will be presented such that radiologists must read a mixture of cases without or with ML support during each round of reading including a wash out period. We will have unavoidable incorporation bias, as the expert reference panel will use the MRI as part of the reference standard. The reference panel will consist of a single person's opinion, which is a limitation to our study. If resources had allowed, the gold standard would have been to have two blinded opinions with a consensus panel in cases of disagreement. Other limitations include varying scan quality as data are acquired over a 9-year period; and replicating clinical reporting in a retrospective study setting can be challenging.

In conducting this study, we will have acquired possibly the largest set of characterised myeloma patient MRI scans in the UK and we anticipate that this will form the basis of a unique training resource in the future.

ML techniques in WB-MRI scans of patients with myeloma are likely to be transferable to other malignancies. In prostate and breast cancer, quantification of metastatic bone disease is an unmet need as bone only disease is not uncommon and is currently classified as non-measurable by RECIST V.1.1.²³ The participating HVs will be consented to allow the anonymised datasets to be a future resource for the wider research community.

Study status

The MALIMAR study opened on 26 April 2018 using protocol V.1.0 (30 October 2017). The study was in phase II, using protocol V.3.0 (31 January 2019), at date of submission. Protocol amendments are documented in online supplemental file 4.

Author affiliations

¹Royal Marsden Hospital NHS Trust, London, UK

²Department of Cancer and Surgery, Imperial College London, London, UK

³Institute of Cancer Research, London, UK

⁴Department of Computing, Imperial College London, London, UK

⁵Department of Radiology, Imperial College Healthcare NHS Trust, London, UK

⁶King's College London, London, UK

Acknowledgements We acknowledge NHS funding to the NIHR Biomedical Research Centre (BRC) at The Royal Marsden and Institute of Cancer Research and the NIHR Royal Marsden Clinical Research Facility. We acknowledge the support of the Imperial College London NIHR BRC Imaging Theme and the Cancer Research UK (CRUK) Imperial Centre and the Imaging Research Office at ICHT. We acknowledge

the support of the CRUK funded National Cancer Imaging Translational Accelerator award (Institute of Cancer Research and Imperial College London).

Contributors AR, CM, TaB, BG, SW, TQ, ThB, SD, MOL, MK and DK: conceptualisation and methodology; AR, CM, TaB, ThB, MK, BG, TQ, XF and SW: investigation; EA and AR: resources; ThB and SD: data curation; LS and EG: formal analysis; AR, DK and CM: supervision; LS: writing—original draft; AR, CM, TaB, BG, LW and LS: writing—review and editing; BG, TQ, XF and SW: data visualisation; LW project administration; EA and AR: funding acquisition.

Funding This study (ID: 16/68/34) is funded by the Efficacy and Mechanism Evaluation (EME) Programme, an MRC and NIHR partnership. In addition, the Department of Radiology has agreed to fund the cost of healthy volunteer whole body MRI scans. The cost of recruitment and consenting of healthy volunteers will be requested through the NHS Clinical Research Network. The views expressed in this publication are those of the authors and not necessarily those of the MRC, NHS, the NIHR, or the Department of Health and Social Care. EG and LS's posts are part funded by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and the Institute of Cancer Research, London. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. SW is supported by the UKRI London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare.

Competing interests AR receives honoraria for educational lecture at Garmisch International Symposium, has an unpaid role on the European Society of Radiology Board of Directors and receives travel cost support where necessary. BG receives grants from other entities; EU commission and UKRI London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare, is a Scientific advisor for Kheiron Medical Technologies (January 2018–September 2021) and receives stock options as part of standard employment packages from both Kheiron Medical Technologies and HeartFlow. EA has a patent pending for Machine Learning in Alzheimer's disease and has a role on the scientific advisory board for Radiopharm Theranostics Limited. MK receives grants from both Myeloma UK and Celgene/BMS, and consulting fees or payments from AbbVie, BMS/Celgene, Janssen, GSK, Karyopharm, Takeda and Seagen. CM & DK receive additional funding as a co-investigator on a radiology NIHR study and is part of the joint venture Celescan with the Royal Marsden, The Institute of Cancer Research and Sopra Steria. TB receives additional funding from CRUK grant funding (NCITA) and NIHR (HTA) and receives honoraria from Bayer.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; peer reviewed for ethical and funding approval prior to submission.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Laura Satchwell <http://orcid.org/0000-0002-2935-6532>

REFERENCES

- 1 NICE. Myeloma: diagnosis and management NICE guideline [NG35], 2016
- 2 Rajkumar SV, Dimopoulos MA, Palumbo A, *et al*. International myeloma Working group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol* 2014;15:e538–48.

- 3 Dimopoulos MA, Hillengass J, Usmani S, *et al.* Role of magnetic resonance imaging in the management of patients with multiple myeloma: a consensus statement. *J Clin Oncol* 2015;33:657–64.
- 4 Pearce T, Philip S, Brown J, *et al.* Bone metastases from prostate, breast and multiple myeloma: differences in lesion conspicuity at short-tau inversion recovery and diffusion-weighted MRI. *Br J Radiol* 2012;85:1102–6.
- 5 Squillaci E, Manenti G, Di Stefano F, *et al.* Diffusion-Weighted MR imaging in the evaluation of renal tumours. *J Exp Clin Cancer Res* 2004;23:39–46.
- 6 Dutoit JC, Vanderkerken MA, Anthonissen J, *et al.* The diagnostic value of Se MRI and DWI of the spine in patients with monoclonal gammopathy of undetermined significance, smouldering myeloma and multiple myeloma. *Eur Radiol* 2014;24:2754–65.
- 7 Messiou C, Hillengass J, Delorme S, *et al.* Guidelines for acquisition, interpretation, and reporting of whole-body MRI in myeloma: myeloma response assessment and diagnosis system (MY-RADS). *Radiology* 2019;291:5–13.
- 8 Wu L-M, Gu H-Y, Zheng J, *et al.* Diagnostic value of whole-body magnetic resonance imaging for bone metastases: a systematic review and meta-analysis. *J Magn Reson Imaging* 2011;34:128–35.
- 9 Messiou C, Porta N, Sharma B, *et al.* Prospective evaluation of whole-body MRI versus FDG PET/CT for lesion detection in participants with myeloma. *Radiol Imaging Cancer* 2021;3:e210048.
- 10 Pawlyn C, Fowkes L, Otero S, *et al.* Whole-Body diffusion-weighted MRI: a new gold standard for assessing disease burden in patients with multiple myeloma? *Leukemia* 2016;30:1446–8.
- 11 Mateos M-V, Hernández M-T, Giraldo P, *et al.* Lenalidomide plus dexamethasone for high-risk smouldering multiple myeloma. *N Engl J Med* 2013;369:438–47.
- 12 Padhani AR, Koh D-M, Collins DJ. Whole-Body diffusion-weighted MR imaging in cancer: current status and research directions. *Radiology* 2011;261:700–18.
- 13 The Royal College of Radiologists London. *Radiologists rC of. Clinical radiology UK workforce census 2015 report*, 2016.
- 14 Juntu J, Sijbers J, De Backer S, *et al.* 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;31:680–9.
- 15 Pauly O, Glocker B, Criminisi A. Fast multiple organ detection and localization in whole-body Mr Dixon sequences. in: International Conference on medical image computing and computer-assisted intervention. *Springer* 2011:239–47.
- 16 Lavdas I, Rockall AG, Castelli F, *et al.* Apparent diffusion coefficient of normal abdominal organs and bone marrow from whole-body DWI at 1.5 T: the effect of sex and age. *AJR Am J Roentgenol* 2015;205:242–50.
- 17 Lavdas I, Glocker B, Rueckert D, *et al.* Machine learning in whole-body MRI: experiences and challenges from an applied study using multicentre data. *Clin Radiol* 2019;74:346–56.
- 18 Lavdas I, Glocker B, Kamnitsas K, *et al.* Fully automatic, multiorgan segmentation in normal whole body magnetic resonance imaging (MRI), using classification forests (CFs), convolutional neural networks (CNNs), and a multi-atlas (MA) approach. *Med Phys* 2017;44:5210–20.
- 19 Kaiser MF, Porta N, Sharma B, *et al.* Prospective comparison of whole body MRI and FDG PET/CT for detection of multiple myeloma and correlation with markers of disease burden: results of the iTIMM trial. *JCO* 2021;39:8012.
- 20 Giles SL, deSouza NM, Collins DJ, *et al.* Assessing myeloma bone disease with whole-body diffusion-weighted imaging: comparison with X-ray skeletal survey by region and relationship with laboratory estimates of disease burden. *Clin Radiol* 2015;70:614–21.
- 21 Giles SL, Messiou C, Collins DJ, *et al.* Whole-Body diffusion-weighted MR imaging for assessment of treatment response in myeloma. *Radiology* 2014;271:785–94.
- 22 Valindria VV, Lavdas I, Bai W, *et al.* Reverse classification accuracy: predicting segmentation performance in the absence of ground truth. *IEEE Trans Med Imaging* 2017;36:1597–606.
- 23 Lecouvet FE, Talbot JN, Messiou C, *et al.* Monitoring the response of bone metastases to treatment with magnetic resonance imaging and nuclear medicine techniques: a review and position statement by the European organisation for research and treatment of cancer imaging group. *Eur J Cancer* 2014;50:2519–31.