# Radiopsy: a prospective observational study on quantitative multiparametric whole-body magnetic resonance imaging to discriminate between smoldering and multiple myeloma

Multiple myeloma (MM), the second most common hematologic malignancy, progresses from a premalignant stage termed monoclonal gammopathy of undetermined significance, with smoldering multiple myeloma (SMM) as an intermediate state. While SMM progresses at a rate of ~10% per year, a subset transitions to MM more rapidly (~14-15%/year). The International Myeloma Working Group (IMWG) updated diagnostic criteria to initiate therapy earlier is based on SLiM biomarkers: ≥60% bone marrow plasma cells (BMPC), serum free light chain ratio ≥100, or ≥1 lesion of ≥5 mm detected by magnetic resonance imaging (MRI).[1] The IMWG recommends low-dose whole-body computed tomography as the first-line imaging technique, reserving whole body-magnetic resonance imaging (WB-MRI) for equivocal cases, because of the limited availability of WB-MRI. However, advanced imaging allows detailed assessment of both diffuse and focal plasma cell infiltration.[2,3]

The ability of sensitive imaging techniques to differentiate between discrete areas of diffuse and focal plasma cell infiltration adds an entirely new dimension to assessments of disease burden and treatment efficacy. This study introduces the concept of "radiopsy": the evaluation of quantitative imaging biomarkers (QIB) near biopsy sites using WB-MRI apparent diffusion coefficient (ADC) and relative fat fraction (rFF), excluding focal lesions, to reflect diffuse disease in the pelvic bone as previously proposed by Pawlyn *et al.*[4]

The aim of the current study was to develop predictive models for clinical diagnosis based on radiological features, enabling virtual non-invasive biopsies to assess disease burden and monitor treatment response, even in areas distant from biopsy sites. This single-center prospective study was approved by the local Ethical Committee (Comitato Etico della Romagna [CEROM] protocol code: IRST 100.15 Accu.MRI) and conducted according to the Declaration of Helsinki.

Between January 2021 and January 2024, consecutive patients with suspected SMM or MM were enrolled. Inclusion required a 3T WB-MRI performed according to MY-RADS[5] and diagnostic confirmation via laboratory and biopsy data. Exclusion criteria included MRI-unsafe implants or other malignancies. Further causes of exclusion from analysis included suboptimal or incomplete imaging. Baseline data (clinical presentation, laboratory parameters, and biopsy results) were collected. Bone marrow biopsies were performed at the right posterior iliac crest as per standard practice. Glomerular filtration rate was calculated using Chronic Kidney Disease -Epidemiology Collaboration equations. The patients' characteristics are summarized in *Online Supplementary Table S1*. Continuous variables are reported as mean ± standard deviation or median with interquartile range, categorical variables as counts and percentages. Mann-Whitney and $\chi_2$ tests were used to compare the SMM and MM groups.

All WB-MRI were acquired on a 3T scanner (Ingenia Philips) using standard MY-RADS protocols.[5] Weekly quality control for ADC and rFF was performed using phantoms. Six volumes of interest (VOI) (2.6 cm³ each) were placed on pelvic and vertebral sites (right inferior [RI], right superior [RS], left inferior [LI], left superior [LS], lumbar vertebra [LX], dorsal vertebra [DX]) on both ADC and FF sequences as shown in *Online Supplementary Figure S1*. RS, closest to the biopsy site, was used for feature extraction and correlation with clinical biomarkers via Pearson's ρ. QIB (histogram, second and higher-order) were extracted using IBSI-compliant S-IBEX software. Voxel resampling (1x1x1 mm) and intensity discretization (32 bins) were applied. Features with an intra-class correlation coefficient (ICC) <0.5 between RS and RI were excluded. Highly correlated features (ρ>0.85) were reduced by selecting those with higher ICC.

LASSO logistic regression was used to select the most predictive QIB (for ADC and rFF separately) across all VOI. Selected features were used to build predictive models. Single- and multi-feature models were validated using 3-fold cross-validation repeated 30 times. Model performance was assessed using area under the curve (AUC) of the receiving operating characteristic curve and DeLong test. The best-performing RS-based model was compared against models using BMPC, mean rFF, or mean ADC. Calibration and decision curves were used for further evaluation. The final RS model was tested on distant VOI (RI, LS, LI, LX, DX) to evaluate generalizability.

We included 102 patients; 18 were excluded due to suboptimal image quality or incomplete scans. The final cohort comprised 84 patients (46 males, 38 females) of whom 45 had MM and 39 had SMM. Clinical and biological characteristics are summarized in Table 1 with Mann-Whitney test results.

A total of 144 histogram and texture QIB were extracted from six VOI (ADC and rFF sequences). After applying ICC and Pearson's ρ filtering (threshold = 0.85), we retained 15

independent QIB per modality. LASSO regression identified four frequently selected predictors per sequence: for rFF – mean intensity, complexity, zone distance non-uniformity, and coefficient of variation; for ADC – mean intensity, zone size non-uniformity, 90th percentile, and zone distance non-uniformity.

Figure 1 shows the correlation of mean ADC and FF with the patients' clinical and laboratory characteristics. Logistic regression models built with these features yielded a mean AUC of 0.80 in the training dataset and 0.70 in the test dataset, with no significant AUC differences across VOI (P>0.05, DeLong test).

Figure 2 compares radiopsy (RS VOI) performance with models using only BMPC or mean rFF. Radiopsy showed a median AUC of 0.80 (0.74-0.87) in the training dataset and 0.76 (0.59-0.885) in the test dataset. Mean rFF alone achieved an AUC of 0.74 in the training dataset and 0.72 in the test dataset. The difference was statistically significant in the training set (P<0.01) but not in the test set (P=0.67). The BMPC-only model outperformed both (AUC=0.89-0.90 in the training/test set; P<0.01). Median precision and recall were 0.70 and 0.80, respectively, for radiopsy, 0.62 and 0.80 for rFF, and 0.82 and 0.75 for BMPC.

Calibration (*Online Supplementary Figure S2A*) showed good alignment between predicted and actual outcomes. Decision analysis (*Online Supplementary Figure S2B*) indicated that BMPC offers the best net benefit in high-risk thresholds. The best-performing radiopsy model (mean intensity, complexity, zone distance non-uniformity, and coefficient of variation from rFF) achieved AUC of 0.76 in

the test dataset and was applied to distant VOI. Performance across distance was 0.71, 0.73, 0.66, 0.66, and 0.69. Differences across sites were not statistically significant (P>0.05), supporting the model's robustness.

Our findings support the potential of WB-MRI-derived QIB, particularly ADC and rFF, in distinguishing MM from SMM. The radiopsy models predicted clinical status non-invasively across multiple VOI, maintaining stable performance even when applied to regions distant from the biopsy site. The models relied heavily on mean %FF, consistently reported as a key imaging biomarker. The 90th percentile of ADC intensity, previously associated with early treatment response, was the second most relevant feature. Additional, though less influential, predictors included %FF coefficient of variation and mean ADC, both cited in earlier MM imaging studies.[6,7] The best results were observed for the RS VOI (AUC: 0.80 training set, 0.76 test set), supporting the value of QIB extracted near biopsy sites. Comparable performance across other pelvic VOI (e.g., LS: 0.73; RI: 0.71 test AUC) suggests the model's robustness and potential for broader anatomical application.

When compared with the BMPC part of the SLiM criteria, the radiopsy model had inferior performance (BMPC test AUC=0.90), as expected. However, radiopsy outperformed single-feature models using mean rFF, especially in training (statistically significant), highlighting the added value of higher-order QIB. Although not statistically significant in the test set, radiopsy still outperformed mean %FF (AUC 0.76 *vs.* 0.72), indicating its improved discriminative power. Calibration curves confirmed good agreement with observed

**Table 1.** Patients' demographic and laboratory data.

| Variables | SMM mean | SMM SD | MM mean | MM SD | P |
|---|---|---|---|---|---|
| Age, years | 58.91 | 10.38 | 64.93 | 9.47 | 0.025 |
| Males/females (N/N) | 18/20 | - | 28/18 | - | 0.216 |
| M-protein, g/L | 19.45 | 10.1 | 28.19 | 23.73 | 0.211 |
| BMPC, % | 25.3 | 15.45 | 63.83 | 25.97 | 0 |
| Hemoglobin, g/dL | 12.44 | 1.71 | 10.69 | 2.08 | 0.001 |
| Platelet count, x$10^9$/L | 231.87 | 60.6 | 205.17 | 79.34 | 0.202 |
| Serum calcium, mg/dL | 9.48 | 0.35 | 9.53 | 0.74 | 0.932 |
| Serum albumin, g/L | 42.67 | 2.97 | 38.6 | 5.64 | 0.001 |
| Albumin-corrected calcium, mg/dL | 9.30 | 0.31 | 9.63 | 0.64 | 0.01 |
| LDH, U/L | 160.53 | 30.22 | 167.91 | 53.03 | 0.969 |
| GFR, mL/min | 81.65 | 20.15 | 73.06 | 22.97 | 0.11 |
| $\beta_2$ microglobulin, mg/L | 2.28 | 0.83 | 5.25 | 4.29 | 0 |
| Alkaline phosphatases, U/L | 63.5 | 19.1 | 91.69 | 75.25 | 0.104 |

SMM: smoldering multiple myeloma; MM: multiple myeloma; SD: standard deviation; BMPC: bone marrow plasma cells; LDH: lactate dehydrogenase; GFR: glomerular filtration rate.

outcomes, while decision-curve analysis showed that BMPC remains superior for high-risk prediction. Encouragingly, radiopsy maintained good performance for distant VOI (AUC: 0.75 pelvic, 0.69 vertebrae), indicating potential for assessing diffuse disease beyond biopsy sites. However, the drop in performance in vertebrae suggests that pelvic bone remains a more reliable region for prediction. The main limitations of this study are the single-center design and use of a single MRI scanner, factors which limit the generalizability of the findings. A multicenter validation is ongoing. Another limitation is the absence of histological confirmation for distant VOI given that multiple biopsies
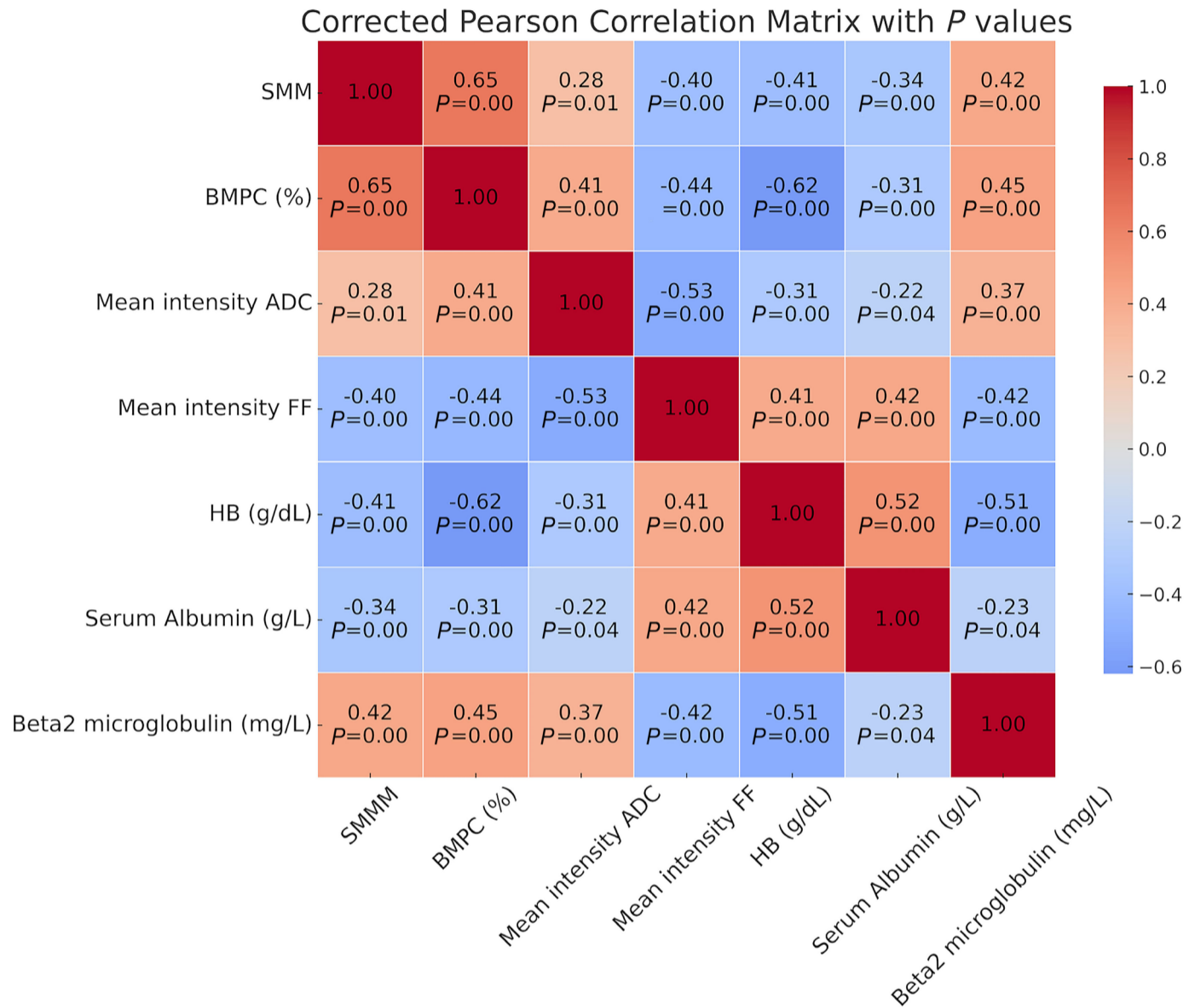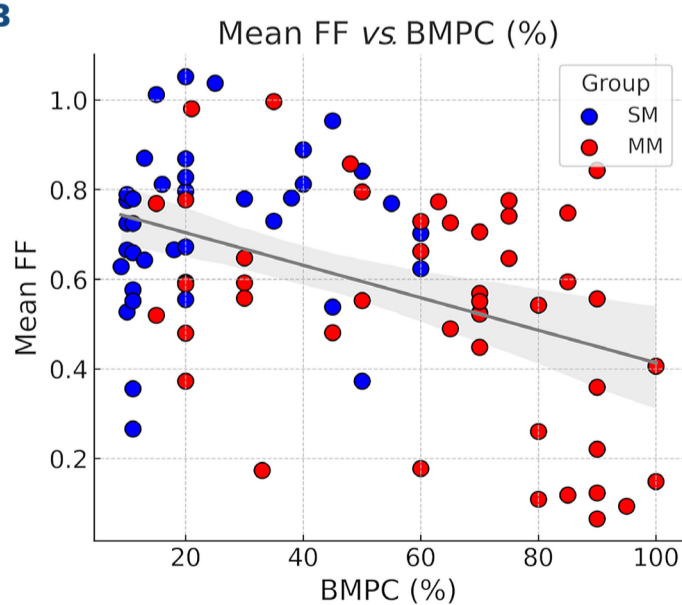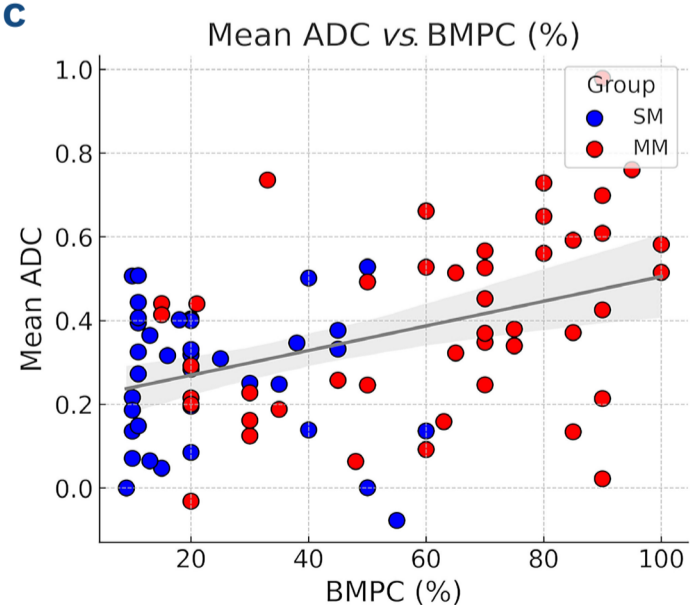


**Figure 1. Correlation between mean intensity of the apparent diffusion coefficient and relative fat fraction and relevant laboratory biomarkers.** (A) Pearson ρ correlation matrix with *P* values. (B, C) Scatterplots with trendlines between bone marrow plasma cell infiltration percentage and relative fat fraction (B) or apparent diffusion coefficient (C). SMM: smoldering multiple myeloma; BMPC: bone marrow plasma cells; ADC: apparent diffusion coefficient; FF: fat fraction; HB: hemoglobin.
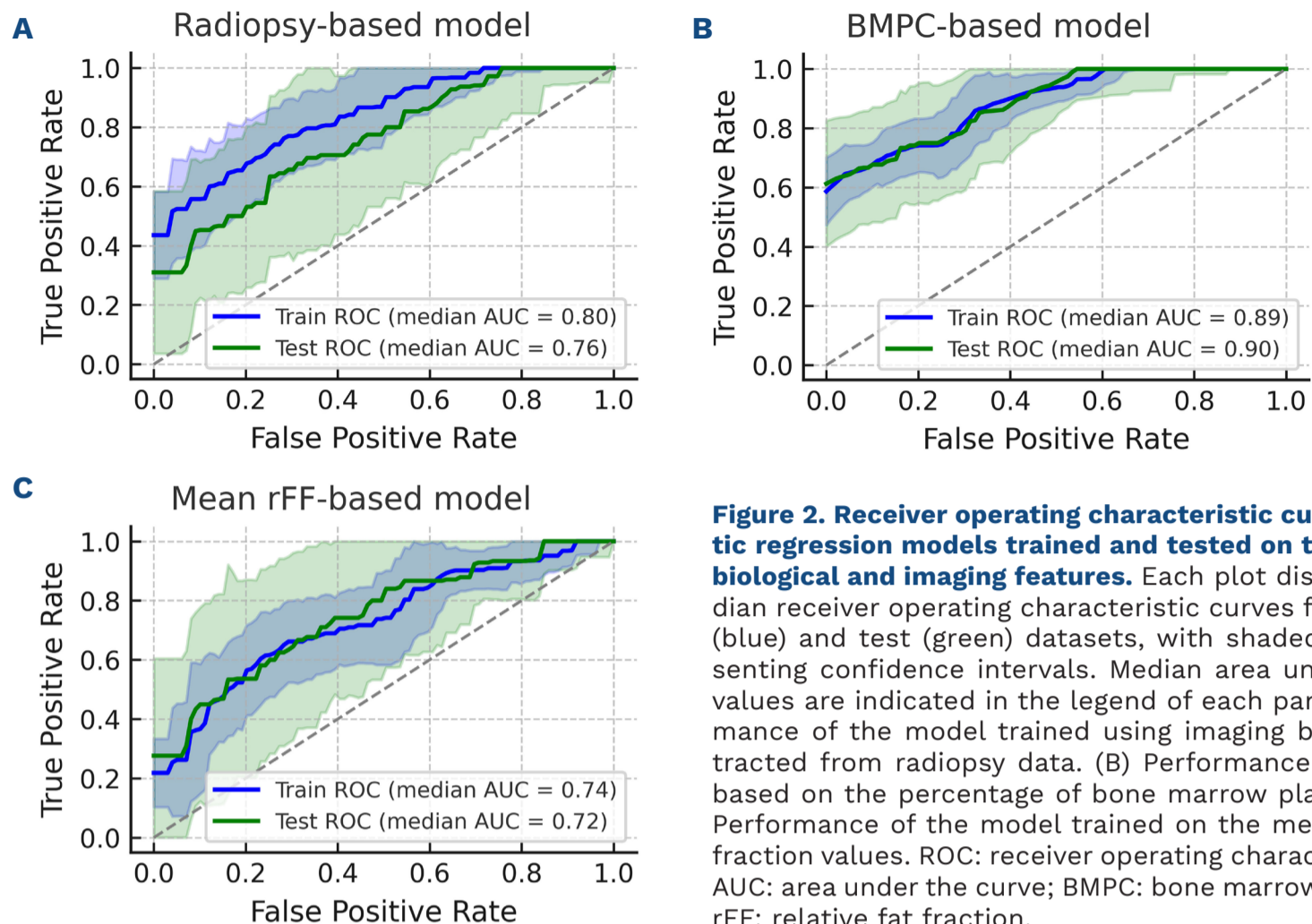
**Figure 2. Receiver operating characteristic curves for logistic regression models trained and tested on three different biological and imaging features.** Each plot displays the median receiver operating characteristic curves for the training (blue) and test (green) datasets, with shaded areas representing confidence intervals. Median area under the curve values are indicated in the legend of each panel. (A) Performance of the model trained using imaging biomarkers extracted from radiopsy data. (B) Performance of the model based on the percentage of bone marrow plasma cells. (C) Performance of the model trained on the mean relative fat fraction values. ROC: receiver operating characteristic curve; AUC: area under the curve; BMPC: bone marrow plasma cells; rFF: relative fat fraction.

were neither ethically nor clinically feasible. This underscores the need for future studies incorporating biopsy-imaging correlations across multiple anatomical regions. Supplementary analyses confirmed the key role of %FF and ADC in predicting BMPC ($\rho=0.44$ and $0.41$, respectively), aligning with the findings of Sun *et al.* ($\rho=0.60$ and $0.49$). Wennmann *et al.*, in a large multicenter study, also found high a correlation ($\rho=0.71$) between QIB and BMPC, although focused on direct numerical comparisons rather than classification. Latifoltojar *et al.* and Koutoulidis *et al.* found high AUC values for rFF features in distinguishing treatment responses.[7,8] Our results confirm the diagnostic relevance of similar QIB, and future work will investigate their role in predicting progression from SMM to MM. Reproducibility remains a key concern. Wennmann *et al.*[9] reported that only 4% of ADC texture QIB achieved acceptable reproducibility across scanners. Our findings support this, with low reproducibility for ADC features (ICC <0.5). In contrast, FF-derived QIB demonstrated better stability. In sum, while ADC-based QIB show promise, their variability limits standalone use. Radiopsy, by combining stable and informative FF-based features, provides a non-invasive and reproducible tool to complement traditional biopsy in disease assessment. Radiopsy differs from conventional imaging biomarkers by integrating multiple radiomic features to capture spatial and textural heterogeneity, beyond simple mean values such as ADC or rFF. This approach supports more refined, non-invasive evaluation of disease burden, including in unbiopsied regions. Radiopsy may also aid in treatment monitoring and evaluation of measurable residual disease, especially when repeat biopsies are not feasible. In conclusion, the integration of QIB with existing clinical biomarkers holds promise for improving the diagnosis and prognosis of MM. Future studies with larger cohorts of patients and external validation are warranted to confirm these findings and further refine the clinical utility of radiopsy in this context.

# Authors

Giacomo Feliciani,[1] Emiliano Loi,[1] Andrea Prochowski Iamurri,[2] Elena Amadori,[2] Eleonora Antognoni,[3] Danila Diano,[2] Daniele Bruno,[1] Arrigo Cattabriga,[4] Paola Caroli,[5] Matteo Marchesini,[6] Davide Bezzi,[7] Delia Cangini,[8] Federica Matteucci,[5,7] Domenico Barone,[2] Davide Nappi,[8] Gerardo Musuraca,[8] Michela Ceccolini,[8] Sonia Ronconi,[8] Irene Azzali,[9] Giovanni Martinelli,[10] Lorenzo Stefano Maffioli,[11] Nicola Normanno,[12] Anna Sarnelli,[1] Alice Rossi[2#] and Claudio Cerchione[8#]

[1]Medical Physics Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola; [2]Radiology Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola; [3]School of specialization in Radiology, University of Ferrara, Ferrara; [4]Department of Radiology, IRCCS Azienda Ospedaliero-Universitaria di Bologna; Dipartimento di Scienze Mediche e Chirurgiche, Bologna; [5]Nuclear Medicine Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola; [6]Translational-Hematology Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino

Amadori", Meldola; [7]Nuclear Medicine Unit, AUSL Romagna, Forlì; [8]Hematology Unit, IRCCS Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST), Meldola; [9]Biostatistics and Clinical Trials Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola; [10]Department of Medical and Surgical Sciences, University of Bologna, Bologna; [11]General Directorate, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola and [12]Scientific Directorate, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy

[#]*AR and CC contributed equally as senior authors.*

Correspondence:
A. ROSSI - alice.rossi@irst.emr.it
C. CERCHIONE - claudio.cerchione@irst.emr.it

**Contributions**
GF designed research, performed research, analyzed and interpreted data, and wrote the manuscript. EL collected data and wrote the manuscript. EAm, EAn, AP, DD, DBr, AC, MM, DBe, DC and MC collected data and performed research. PC and DN designed and performed research, and collected data. FM, DBa, GMu, GMa, LSM, NN and AS designed research. IA performed the statistical analysis and research. AR and CC designed and performed research, and wrote the manuscript.

**Data-sharing statement**
Quantitative imaging biomarkers used to develop the models are available as *Online Supplementary Data*. MRI images and data, QIB and laboratory data of the patients are available upon reasonable request to the corresponding author.

# References

1. Rajkumar SV, Dimopoulos MA, Palumbo A, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. Lancet Oncol. 2014;15(12):e538-e548.
2. Rossi A, Cattabriga A, Bezzi D. Symptomatic myeloma: PET, whole-body MR imaging with diffusion-weighted imaging or both. PET Clin. 2024;19(4):525-534.
3. Zamagni E, Tacchetti P, Cavo M. Imaging in multiple myeloma: how? When? Blood. 2019;133(7):644-651.
4. 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(6):1446-1448.
5. 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(1):5-13.
6. Sun M, Cheng J, Ren C, et al. Evaluation of diffuse bone marrow infiltration pattern in monoclonal plasma cell diseases by quantitative whole-body magnetic resonance imaging. Acad Radiol. 2022;29(4):490-500.
7. Koutoulidis V, Terpos E, Papanikolaou N, et al. Comparison of MRI features of fat fraction and ADC for early treatment response assessment in participants with multiple myeloma. Radiology. 2022;304(1):137-144.
8. Latifoltojar A, Hall-Craggs M, Bainbridge A, et al. Whole-body MRI quantitative biomarkers are associated significantly with treatment response in patients with newly diagnosed symptomatic multiple myeloma following bortezomib induction. Eur Radiol. 2017;27(12):5325-5336.
9. Wennmann M, Thierjung H, Bauer F, et al. Repeatability and reproducibility of ADC measurements and MRI signal intensity measurements of bone marrow in monoclonal plasma cell disorders: a prospective bi-institutional multiscanner, multiprotocol study. Invest Radiol. 2022;57(4):272-281.