ORIGINAL RESEARCH

# Validation of a Diagnostic Model to Differentiate Multiple Myeloma from Bone Metastasis

Phichayut Phinyo [ID][1−4,]*, Nutcha Jarupanich[2], Lalita Lumkul [ID][2,4], Areerak Phanphaisarn[3,5], Thanate Poosiripinyo [ID][6], Sermsak Sukpanichyingyong[6], Pichaya Thanindratarn[7], Yodsawee Pornmeechai[8], Taweechok Wisanuyotin[9], Rapin Phimolsarnti[10], Ekarat Rattarittamrong[11], Dumnoensun Pruksakorn[3−5,]*

[1]Department of Family Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand; [2]Center for Clinical Epidemiology and Clinical Statistics, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand; [3]Musculoskeletal Science and Translational Research (MSTR), Chiang Mai University, Chiang Mai, Thailand; [4]Center of Multidisciplinary Technology for Advanced Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand; [5]Department of Orthopaedics, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand; [6]Department of Orthopaedics, Khon Kaen Hospital, Khon Kaen, Thailand; [7]Chulabhorn Hospital, Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy, Bangkok, Thailand; [8]Department of Orthopedics, Queen Savang Vadhana Memorial Hospital, Chonburi, Thailand; [9]Department of Orthopaedics, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand; [10]Department of Orthopaedic Surgery, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; [11]Division of Hematology, Department of Internal Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

*These authors contributed equally to this work

Correspondence: Phichayut Phinyo, Department of Family Medicine, Faculty of Medicine, Chiang Mai University, 110 Intawaroros Road, Si Phum, Muang, Chiang Mai, 50200, Thailand, Tel +66-5393-6150, Fax +66-5393-6150, Email phichayutphinyo@gmail.com; phichayut.phinyo@cmu.ac.th; Dumnoensun Pruksakorn, Department of Orthopaedics, Faculty of Medicine, Chiang Mai University, 110 Intawaroros Road, Si Phum, Muang, Chiang Mai, 50200, Thailand, Tel +66-5393-6150, Fax +66-5393-6150, Email dumnoensun@hotmail.com; dumnoensun.p@cmu.ac.th

**Purpose:** A diagnostic model to differentiate multiple myeloma (MM) from bone metastasis (BM) in patients with destructive bone lesions (MM-BM DDx) was developed to promote timely and appropriate referral of patients with MM to hematologists. External validation has never been conducted. This study aims to externally validate the performance of the MM-BM DDx model.

**Patients and Methods:** This multi-center external validation study was conducted using retrospective data of patients over 45 years old diagnosed with MM or BM at six university-affiliated hospitals in Thailand from 2016 to 2022. The MM-BM DDx development dataset, including patients from 2012 to 2015, was utilized during external validation. Diagnostic indicators for MM included in the MM-BM DDx model are serum creatinine, serum globulin, and serum alkaline phosphatase (ALP). MM and BM diagnosis was based on the documented International Classification of Diseases 10th Revision codes. Model performance was evaluated in terms of discrimination, calibration, and accuracy.

**Results:** A total of 3018 patients were included in the validation dataset (586 with MM and 2432 with BM). Clinical characteristics were similar between the validation and development datasets. The MM-BM DDx model's predictions showed an AUC of 0.89 (95% CI, 0.87, 0.90). The predicted probabilities of MM from the model increased concordantly with the observed proportion of MM within the validation dataset. The estimated sensitivity, specificity, and LR for each odds class in the validation dataset were similar to those of the development dataset.

**Conclusion:** The discriminative ability and calibration of the MM-BM DDx model were found to be preserved during external validation. These findings provide support for the practical use of the MM-BM DDx model to assist clinicians in identifying patients with destructive bone lesions who are likely to have MM and enable them to arrange timely referrals for further evaluation by hematologists.

**Keywords:** diagnosis, primary care, clinical prediction model, referral, bone metastases, multiple myeloma

## Plain Language Summary

It is often challenging to distinguish between multiple myeloma (MM) and bone metastasis (BM) due to their shared characteristics. Consequently, misdiagnoses and inappropriate referrals are common. To address this issue, a diagnostic model was recently developed to differentiate between these two conditions. In this multi-center validation study, we conducted the first validation of the MM-BM DDx model. The model exhibited preserved performance and robust accuracy, supporting its utilization in clinical practice. By

incorporating the MM-BM DDx model, clinicians can promote appropriate and timely referrals of patients with a high likelihood of MM to hematologists. This approach has the potential to reduce the duration of MM diagnosis and treatment, leading to improved prognosis and survival outcomes.

## Introduction

Multiple myeloma (MM) and secondary bone metastasis (BM) are two primary differential diagnoses for adult patients presenting with bone pain and destructive bone lesions.[1] An accurate initial diagnosis is crucial, as these conditions have major distinctions in terms of disease prognosis, subsequent investigations, and therapeutic goals.[2] However, in practice, it is often challenging to differentiate between these two conditions due to their similar clinical and radiographic characteristics.[3–5] Consequently, misdiagnosis at the first medical contact and inappropriate referrals are common.[6,7] Such occurrences have a significant impact, particularly on MM patients, as delayed myeloma diagnosis has been associated with substantial morbidity and worsened survival.[6] Considering the continuous rise in incidence rates of both MM and BM in recent years,[8,9] the number of patients presenting with destructive bone lesions is expected to increase. This poses a significant challenge for primary care physicians who are typically the first to encounter such patients. Therefore, there is a need for a clinical diagnostic tool that can estimate the likelihood of MM and guide physicians in determining whether referral to hematologists is necessary.[7]

In 2020, a diagnostic tool to differentiate MM from BM in patients with destructive bone lesions (MM-BM DDx) was developed in a large series of patients diagnosed with MM and BM in northern Thailand.[10,11] The model includes three simple, routinely available laboratory parameters: serum creatinine, serum globulin, and serum alkaline phosphatase (ALP). The model demonstrated excellent discriminative ability and calibration. However, its reproducibility and applicability outside the original dataset were still unknown. Therefore, the objective of this study was to externally validate the performance of the MM-BM DDx model using a multi-center validation dataset that included patients from six tertiary care centers across Thailand.

## Materials and Methods
### Source of Data and Participants

This external validation study is a multi-center, retrospective cross-sectional study. Six university-affiliated hospitals and medical schools participated in this study: Faculty of Medicine, Chiang Mai University, Khon Kaen University, Siriraj Hospital at Mahidol University, Queen Savang Vadhana Memorial Hospital, Khon Kaen Hospital, and Chulabhorn Hospital. The development dataset of the MM-BM DDx model,[10] including patients diagnosed with either MM or BM at Chiang Mai University Hospital from 2012 to 2015, was also utilized during this external validation to help interpret the findings.[12]

The Institutional Review Board of each center approved the study protocol (FAM-2564-07951) and waived the requirement of informed consent due to the retrospective nature of data collection. This study was conducted in accordance with the Declaration of Helsinki.[13] We followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statements for reporting.[14]

The intended patient domain for the MM-BM DDx model is an adult patient presenting with bone pain and destructive bone lesions on plain radiographs, whose differential diagnosis is presumably either MM or BM. The inclusion criteria were as follows: (1) patients older than 45 years old and (2) patients who were diagnosed with MM (International Classification of Diseases codes 10th revision (ICD10)-C90) or secondary BM (ICD10-C795). The decision to include only patients aged 45 years and older was based on the findings of the development study of the MM-BM DDx model.[10] This study identified the specific age groups in which secondary BM and MM are the primary differential diagnoses for patients presenting with destructive bone lesions. Records with missing data were included. Each participating center retrieved the patient data between January 1, 2016, and December 31, 2022.

### The MM-BM DDx Model

The MM-BM DDx model incorporates three clinical laboratory values: serum creatinine, serum globulin, and serum ALP. All three laboratory values undergo logarithmic transformation and are inputted into the MM-BM DDx fractional

polynomial logit equation. The linear predictors (LP) of the model are summed to generate predicted log-odds, which are then converted to odds by exponentiation. The likelihood ratio (LR) of being MM is estimated by dividing the predicted odds of MM by the odds of MM in the development dataset, which was 0.30. The detailed equation for predicting the odds of MM is provided in Table 1. The online web application for the MM-BM DDx model is accessible through the following link: https://tinyurl.com/mm-bm-ddx.

The MM-BM DDx model does not incorporate radiographic pattern as one of its predictors due to the high potential for misclassification. Moreover, during the development phase, the inclusion of all patients, irrespective of their radiographic patterns, aimed to broaden the scope of the patient population and include those who were suspected by physicians to have either MM or BM, without relying on precise classification based on radiographic pattern.[10] This approach was adopted to encourage the utilization of the MM-BM DDx model regardless of the radiographic pattern, taking into account that patients with MM can still exhibit osteoblastic or mixed patterns in their imaging findings.

## Outcome
The diagnosis of MM and BM was based on the ICD10 codes (ICD10-C90 for MM and ICD10-C795 for BM). The investigators verified all retrieved records. MM diagnosis was verified according to the standard diagnostic criteria by the International Myeloma Working Group (IMWG).[15]

## Data Acquisition and Definition
We used standardized, de-identified case record form for data collection. The data on patient demographics (ie, age and sex), bone radiographic patterns from the plain film (ie, osteolytic, osteoblastic, and mixed lytic-blastic lesion), and clinical laboratory values (ie, hematologic parameters (ie, hemoglobin, hematocrit), renal function test and electrolytes (ie, blood urea nitrogen, serum creatinine, serum calcium), liver function test (ie, serum protein profiles, and relevant enzyme levels)) were collected at the date of MM or BM diagnosis (or the date closest to the diagnostic date but before the date of definitive treatment). For patients with MM, we collected data on immunoelectrophoretic paraprotein type and serum β2 microglobulin for staging patients according to the International Staging System (ISS). The data on the primary cancer site were collected for patients with secondary BM. Finally, we extracted the patient characteristics and the performance indices of the MM-BM DDx from the development study[10] for comparative purposes.

## Statistical Analysis
Statistical analysis was performed using Stata 17 (StataCorp, USA). Assuming an expected discriminative ability at an area under the receiver operating characteristic curve (AUC) of 0.8 with an error margin of 0.1 and the MM:BM ratio of 1:3, a sample size of 1768 with at least 584 patients with MM was deemed sufficient according to calculation methods proposed by Riley et al[16] Continuous data were presented as mean and standard deviation (SD) or median and

**Table 1** Detailed MM-BM DDx Fractional Polynomial Logit Equation for Predicting Odds of Multiple Myeloma (MM)

| MM-BM DDx | Label | Covariate Transformation | Coefficient |
|---|---|---|---|
| **Predictors** | | | |
| Serum creatinine | CR | Log(CR)-0.0237 | 1.28 |
| Serum globulin | GLOB | Log(GLOB)$^{-0.5}$–0.8714 | −92.64 |
| | | Log(GLOB)$^{-0.5}$×Log (Log(GLOB))-0.2400 | −48.14 |
| Serum ALP | ALP | Log(ALP)-4.9318 | −0.97 |
| Intercept | | | −2.28 |
| Log odds of MM (LP) | −2.28 + (1.28×Log(CR)-0.0237) + (−92.64×Log(GLOB)$^{-0.5}$–0.8714) + (−48.14×Log(GLOB)$^{-0.5}$×Log(Log(GLOB))-0.2400) + (−0.97×Log(ALP)-4.9318) | | |
| Odds of MM | $e^{LP}$ | | |

**Abbreviations**: ALP, alkaline phosphatase; e, a mathematical constant e; LP, linear predictors; Log, logarithm function.

interquartile range (IQR) according to their distribution. Categorical data were presented with frequency and percentage. Fisher's exact test was used for the comparison of categorical data. An independent *t*-test and Mann–Whitney *U*-test were used to compare continuous data as appropriate. All P-values were two-sided and regarded as statistically significant at less than 0.05.

Records with missing data were included in the analysis to avoid selection bias and improve statistical efficiency. Multiple imputation with chain equation (MICE) and predictive mean matching (PMM) were used to impute missing predictor values.[17] All predictors within the MM-BM DDx model together with the final diagnosis were included within the MICE model. Age, sex, hemoglobin level, radiographic patterns of the lesions, and centers were also included as auxiliary variables. We created 30 imputed datasets and combined their estimates using Rubin's rules.[18]

We followed a three-step framework for external validation, recently proposed by Debray et al[12] First, the relatedness between the development and validation datasets was evaluated by estimating the discriminative ability of the membership model, which is a binary logistic model with the dataset (either development or validation) as a dependent variable and predictors of the MM-BM DDx model and MM diagnosis as independent variables. The discriminative ability of the model was expressed using AUC. We also estimated the LP of the MM-BM DDx model and its SD for both datasets, which were used to reflect the distribution of case mix in each dataset. Independent *t*-test and variance-comparison test were used to compare the mean and SD of the LP, respectively. Second, we evaluated the external model performance in three aspects: model discrimination, model calibration, and diagnostic accuracy. For each patient, we used the MM-BM DDx model to predict the odds of MM. The discriminative ability of the model's predicted odds was quantified using AUC. For calibration, all predicted odds were divided into ten groups based on the reported cutoff values from the MM-BM development study.[10] The predicted probability of MM was estimated for each group and plotted against the observed proportion of MM to create calibration plots. The expected:observed (E:O) ratio, calibration-in-the-large (CITL), and calibration slope were estimated. Subgroup analyses were conducted to examine the discriminative ability and model calibration by center. To examine the diagnostic accuracy, we estimated each group's sensitivity, specificity, and LR of the MM diagnosis. Sensitivity and specificity were estimated using a binary logistic model, whereas LR was estimated by dividing the individual's predicted odds by the odds of MM in the development dataset. Third, we interpreted the model validation results based on the first and second steps.

A post-hoc analysis was conducted to assess the performance of the MM-BM DDx model in distinguishing patients with MM from those with secondary BM originating from various primary cancers. Additionally, we planned to investigate the prognostic significance of the MM-BM DDx model for both MM and BM patients. This would involve examining the association between the log-likelihood ratio (logLR) of MM and the International Staging System (ISS), which is a standard prognostic staging system for MM patients,[19] and the association between the logLR for MM and serum ALP tertiles for BM patients. One-way analysis of variance (ANOVA) was used to test the statistical difference of logLR across the three groups for both objectives.

## Results
### Participants
A total of 3018 patients diagnosed with MM or BM were included in the validation dataset (586 with MM and 2432 with BM). Overall, the mean age was 62.8 (SD 10.0), with a male-to-female ratio of 54:46 in the validation dataset. Compared with patients in the development dataset, the clinical characteristics of those in validation datasets were similar, except for the distribution of radiographic patterns and the availability of plain film (Table 2). The proportion of missing data for each characteristic and MM-BM DDx predictor are also shown in Table 2. In the validation datasets, there were 2378 (78.8%) patients with complete data on all of the MM-BM DDx model parameters. The distributions of the three parameters, categorized by diagnosis and datasets, are visualized using box plots (Figure S1).

The majority of patients with MM had either ISS stage II (14.9%) or stage III (47.8%) (Table S1). The most common immunoelectrophoretic paraprotein subtype was IgG (46.8%), followed by IgA (15.5%), and light chain (11.1%) (Table S1). The three commonly reported primary sites of patients with secondary BM were the lung (28.3%), breast (16.5%), and prostate (15.8%) (Table S2).

**Table 2** Clinical Characteristics of the Patients in the Validation Dataset and the Development Dataset

| Clinical Characteristic | Validation Dataset (n=3018) | | | Development Dataset (n=586) | | |
|---|---|---|---|---|---|---|
| | Missing Data | Multiple Myeloma (n=586) | Bone Metastasis (n=2432) | Missing Data | Multiple Myeloma (n=136) | Bone Metastasis (n=450) |
| | n (%) | Mean±SD | Mean±SD | n (%) | Mean±SD | Mean±SD |
| **Demographic characteristics** | | | | | | |
| Age at presentation (years) | 0 (0) | 64.0±9.6 | 62.5±10.0 | 0 (0) | 62.4±9.2 | 63.2±9.9 |
| Sex, (n, %) | | | | | | |
| Male | 0 (0) | 309 (52.7) | 1307 (53.7) | 0 (0) | 75 (55.2) | 253 (56.2) |
| **Abnormal bone radiographic pattern from plain film** | | | | | | |
| No plain film (n, %) | 0 (0) | 105 (17.9) | 997 (41.0) | 0 (0) | 47 (34.6) | 204 (45.3) |
| Osteolytic lesion (n, %) | 0 (0) | 477 (81.4) | 1079 (44.4) | 0 (0) | 89 (65.4) | 146 (32.4) |
| Osteoblastic lesion (n, %) | 0 (0) | 1 (0.2) | 222 (9.1) | 0 (0) | 0 (0) | 40 (8.9) |
| Mixed lytic-blastic lesion (n, %) | 0 (0) | 3 (0.5) | 134 (5.5) | 0 (0) | 0 (0) | 60 (13.3) |
| **Clinical Laboratory values** | | | | | | |
| Hemoglobin (g/dL) | 286 (9.5) | 8.8±2.3 | 10.9±2.4 | 39 (6.7) | 9.4±2.4 | 11.2±2.2 |
| Hematocrit (mg %) | 290 (9.6) | 26.7±6.6 | 33.4±6.7 | 44 (7.5) | 28.9±7.5 | 34.5±6.4 |
| BUN (mg/dL, median (IQR)) | 453 (15.0) | 21 (13, 34) | 14 (10, 19) | 52 (8.9) | 19 (12, 30) | 14 (11, 21) |
| Serum creatinine (mg/dL, median (IQR)) | 275 (9.1) | 1.2 (0.9, 2.2) | 0.8 (0.6, 1.0) | 44 (7.5) | 1.3 (0.8, 2.3) | 0.9 (0.7, 1.2) |
| Serum calcium (mg/dL) | 878 (29.1) | 9.8±1.3 | 9.2±1.3 | 146 (24.9) | 9.8±1.6 | 9.2±1.3 |
| Total protein (g/dL) | 579 (19.2) | 9.1±2.5 | 7.2±1.0 | 66 (11.3) | 8.8±2.6 | 7.0±1.0 |
| Serum albumin (g/dL) | 482 (16.0) | 3.2±0.8 | 3.6±0.7 | 57 (9.7) | 3.1±0.9 | 3.4±1.0 |
| Serum globulin (g/dL) | 588 (19.5) | 5.9±2.9 | 3.6±0.8 | 63 (10.8) | 5.7±3.1 | 3.6±1.0 |
| ALP (U/L, median (IQR)) | 568 (18.8) | 79 (63, 111) | 130 (89, 235) | 72 (12.3) | 86 (63, 120) | 138 (96, 262) |
| LDH (U/L, median (IQR)) | 2537 (84.1) | 203.5 (152, 292) | 299 (207, 469) | 511 (86.9) | 201 (142, 245) | 239 (178, 347) |

**Abbreviations**: ALP, alkaline phosphatase; BUN, blood urea nitrogen; IQR, interquartile range; LDH, lactate dehydrogenase; SD, standard deviation.

## Model Performance

Regarding the relatedness between datasets, the membership model showed an AUC of 0.56 with a 95% confidence interval (CI) between 0.53 and 0.58. Both datasets also had a similar mean of the LP (development −1.67 vs validation −1.82, p = 0.135). However, the SDs of the LP were significantly different (development 2.11 vs validation 2.50, p < 0.001).

For model discrimination, the MM-BM DDx model illustrated an AUC of 0.89 (95% CI 0.87, 0.90) in the validation dataset (Figure 1a), which was similar to that of development study at 0.90 (95% CI 0.86, 0.93). For model calibration, the average predicted probabilities of MM were similar to the observed proportions of MM (E:O ratio 1.05, CITL −0.103, and calibration slope 0.97). According to the calibration plot, the predicted probabilities and the observed proportions of MM increased concordantly with one another (Figure 1b). Subgroup analysis results by center for discriminative ability and calibration are presented in Figures S2 and S3.

Table 3 presents the sensitivity, specificity, and LR for each decile of odds in the validation and development dataset. The group-specific sensitivity, specificity, and LR in the validation dataset were similar to those of development study.

In the post-hoc analysis, we focused only on the four most prevalent types of primary cancer sites for BM. The AUC values of the MM-BM DDx model for differentiating various cancers from MM were as follows: 0.90 (95% CI 0.88, 0.92) for lung cancer, 0.91 (95% CI 0.90, 0.93) for breast cancer, 0.87 (95% CI 0.85, 0.90) for prostate cancer, and 0.90 (95% CI 0.88, 0.92) for liver cancer. Among MM patients with available ISS data (n = 394), there was a significant association between the higher predicted logLR of MM and a higher ISS stage (P < 0.001) (Table S3 and Figure S4). In contrast, the lower logLR of MM showed a significant association with higher serum ALP tertiles in patients with BM (n = 1870) (P < 0.001) (Table S3 and Figure S5).
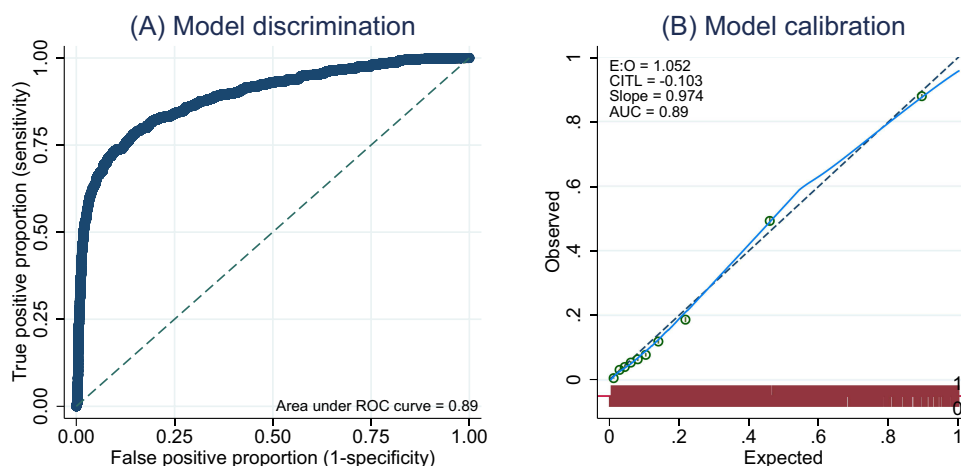
**(A) Model discrimination**

**(B) Model calibration**

**Figure 1** The diagnostic performance of the MM-BM DDx model in the external validation dataset. (**a**) Model discriminative ability illustrated with area under the receiver operating characteristics curve (ROC). (**b**) Model calibration plots illustrating the agreement between the expected probability from the MM-BM DDx model and the observed probability of multiple myeloma in the external validation dataset.

## Discussion

Our study validated the MM-BM DDx model using an independent dataset comprising patients from six centers across Thailand. The MM-BM DDx model demonstrated a discriminative ability similar to that observed during the development phase. The predicted probabilities of MM also closely matched the observed proportion of MM, indicating that the MM-BM DDx model remained well calibrated even when applied to an external dataset. Additionally, when the predicted odds were categorized based on the reported ranges in the development study, the sensitivity, specificity, and LR values for each group were consistent with those reported in the development study.

The similarity in performance between the validation and development datasets can be attributed to their relatedness. According to the poor discriminative ability of the membership model (AUC 0.56), both datasets were strongly related and difficult to distinguish.[12] This was also reflected through the similarity in the mean LP and the outcome proportions between the two datasets. For this reason, it is unlikely that the model's external calibration-in-the-large and calibration would be significantly affected.[12,20] As for the SD of the LP, datasets with larger SD should presumably carry higher discriminative ability than those with lower SD.[12,20] Although the SD of the LP in the validation dataset was significantly higher than that of the development dataset, the discriminative ability was almost equal. This could potentially be attributed to the influence of extreme globulin values observed in the validation dataset.

**Table 3** Diagnostic Accuracy of the MM-BM DDx Model by Predicted Odds Deciles in External Validation and Development Dataset

| Decile | Range of Predicted Odds | Validation Dataset (n=3018) | | | Development Dataset (n=586) | | |
|--------|----------|----------|----------|----------|----------|----------|----------|
| | | Sensitivity (95% CI) | Specificity (95% CI) | Median LR (P2.5, P97.5) | Sensitivity (95% CI) | Specificity (95% CI) | LR (95% CI) |
| D1 | ≤0.0269 | 100 (NE) | 0 (NE) | 0.05 (0.01, 0.09) | 100 (97.3, 100) | 0 (0, 0.02) | 0.06 (0.00, 0.34) |
| D2 | 0.0270–0.0459 | 99.5 (99.0, 100) | 15.5 (13.9, 17.1) | 0.12 (0.09, 0.15) | 99.3 (96.0, 100) | 12.7 (9.7, 16.1) | 0.11 (0.01, 0.45) |
| D3 | 0.0460–0.0669 | 97.0 (95.5, 98.6) | 29.8 (27.8, 31.8) | 0.19 (0.15, 0.22) | 97.8 (93.7, 99.5) | 25.3 (21.4, 29.6) | 0.11 (0.01, 0.46) |
| D4 | 0.0670–0.0869 | 94.6 (92.7, 96.6) | 42.9 (40.8, 45.0) | 0.26 (0.22, 0.29) | 96.3 (91.6, 98.8) | 37.8 (33.3, 42.4) | 0.24 (0.06, 0.67) |
| D5 | 0.0870–0.1179 | 92.3 (89.9, 94.6) | 53.0 (50.9, 55.2) | 0.34 (0.29, 0.39) | 93.4 (87.8, 96.9) | 50.0 (45.3, 54.7) | 0.18 (0.04, 0.56) |
| D6 | 0.1180–0.1499 | 88.0 (85.2, 90.9) | 66.2 (64.1, 68.4) | 0.44 (0.39, 0.49) | 91.2 (85.1, 95.4) | 62.4 (57.8, 66.9) | 0.38 (0.13, 0.92) |
| D7 | 0.1500–0.2269 | 84.4 (81.3, 87.5) | 74.2 (72.2, 76.2) | 0.59 (0.50, 0.74) | 86.8 (79.9, 92.0) | 74.0 (69.7, 78.0) | 0.68 (0.30, 1.40) |
| D8 | 0.2270–0.5259 | 78.7 (75.2, 82.1) | 84.6 (83.0, 86.3) | 1.06 (0.77, 1.70) | 79.4 (71.6, 85.9) | 84.9 (81.2, 88.1) | 1.49 (0.78, 2.76) |
| D9 | 0.5260–3.2219 | 67.0 (63.0, 71.0) | 94.0 (93.0, 95.1) | 3.67 (1.80, 9.86) | 66.2 (57.6, 74.1) | 93.8 (91.1, 95.8) | 4.83 (2.68, 8.77) |
| D10 | ≥3.2220 | 38.3 (34.2, 42.4) | 98.8 (98.3, 99.3) | 62.66 (11.11, 1,903,818) | 40.4 (32.1, 49.2) | 99.1 (97.7, 99.8) | 45.50 (16.26, 174.85) |

**Abbreviations**: CI, confidence interval; D, decile; LR, likelihood ratio; NE, not estimable; P2.5, 2.5th percentile; P97.5, 97.5th percentile.

Overall, we recommend implementing the MM-BM DDx model in clinical practice, particularly in populations similar to our study base, such as Thais or Asian populations. The model should be used for patients presenting with differential diagnoses of either MM or secondary BM. In clinical practice, distinguishing between MM and BM can be challenging, even with advanced imaging techniques.[21,22] In such cases, the MM-BM DDx model can play a crucial role by providing attending physicians with an estimated likelihood of MM based on just three routinely available laboratory values. For clinical applicability, the MM-BM DDx model provided users with the predicted LR of MM instead of odds. The clinical suggestion of the MM-BM DDx model was based on the magnitude and direction of the patient's predicted LR.[23] Patients with LR above 2.0 should be referred to hematologists for further evaluation. In contrast, patients with LR less than 0.5 may not require referral as secondary BM is more likely than MM. Patients whose LR was higher than 0.5 but less than 2.0 should be considered on a case-by-case basis, and referral to hematologists may not be necessary.

As two of the three predictors in the MM-BM DDx model, namely serum creatinine and serum globulin, are established prognostic factors for survival and treatment outcomes in MM patients,[24–28] the model predictions may also have predictive value for these outcomes. However, since follow-up data was not collected in this study, we were unable to directly examine the prognostic significance of our predictions. To indirectly assess the potential prognostic utility of the model, we investigated the association between the higher predicted log-likelihood ratio (logLR) of MM and a higher ISS. This association was straightforward, as patients with a high predicted logLR of MM typically have elevated serum creatinine and globulin levels at diagnosis, which are generally indicative of unfavorable survival outcomes. The MM-BM DDx model also incorporates serum ALP, which is a potential prognostic marker in BM patients. Studies have shown that higher serum ALP levels at BM diagnosis are associated with a higher preoperative metastatic tumor burden, increased osteolytic activity, and serve as a prognostic marker for one-year survival in both spinal and non-spinal metastatic diseases.[29,30] Given that patients with high serum ALP have a low logLR of MM in our model, it can be assumed that BM patients with very low levels of logLR of MM may have poorer survival compared to those with higher logLR of MM. It is important to note that these findings are indirect due to the absence of direct follow-up data in our study. Implications of the model and further longitudinal studies are required to provide direct answers.

There are some major strengths to our study. This study is the first to externally validate the MM-BM DDx model using a large multi-center cross-sectional data of patients who presumably presented with bone pain and destructive bone lesions diagnosed with MM and BM across Thailand. The total study size and the number of events were also achieved as planned and significantly larger than that of the development study. To date, the MM-BM DDx model remains the only multivariable clinical prediction model developed to differentiate adult patients with destructive bone lesions. It is essential to recognize the distinction in the predictive context compared to models or studies aiming to use laboratory panels or clinical predictors for MM screening in a primary care setting,[31,32] where patients are often asymptomatic or have not yet developed bone lesions. While there have been some studies conducted to address similar questions, those studies primarily relied on advanced imaging techniques.[4,5,21,22] In 2021, Annibali et al investigated the diagnostic value of serum ALP in differentiating between BM and MM in adult patients with bone lesions.[33] Their findings demonstrated that a low or normal value of serum ALP was highly predictive of MM, which aligns with the results from our validation dataset.

## Limitations

First, all the data used were retrospectively collected and subjected to bias. However, as all of the predictors used in the MM-BM DDx model were objective laboratory values, the degree of bias should be non-trivial. Second, the data on the MM-BM DDx predictors were missing in 21% of the patients. To maximize statistical efficiency, we performed a MICE procedure to impute missing data assuming the data were missing at random. Third, some participating centers collected data on only a portion of available patients during the study period. Fourth, the cross-sectional study base was not population-analog or consecutive. Without problem-oriented medical records,[34] including the intended study domain who presented with bone pain and destructive bone lesions on imaging was difficult. Moreover, not all patients with MM would present with bone pain and destructive bone lesion, and some patients with BM may be incidentally diagnosed. Thus, we had to assume that all patients diagnosed with MM or BM shared similar clinical presentations, which we believed was reasonable as these conditions were common differential diagnoses for adult patients with destructive bone lesions.[1] In addition, most patients with BM were symptomatic,[35] and 50 to 70% of patients with MM presented with bone pain and destructive bone lesions.[36,37] Other studies

exploring features to differentiate MM from BM also directly contrasted patients diagnosed with MM and BM.[21,33] Fifth, the included study populations were patients in tertiary care centers, which was inconsistent with the primary care patients' intended domain. Finally, despite the robust performance of the MM-BM DDx model during this multi-center validation, its generalizability to non-Asian populations or individuals younger than 45 years old remains unknown. Furthermore, the potential clinical impact of its application in practice is yet to be determined. Further research should be conducted to investigate the clinical outcomes of patients after the model's implementation.

## Conclusion

The MM-BM DDx model has undergone its first external validation as a diagnostic tool for distinguishing MM from secondary BM in patients with destructive bone lesions. With its excellent discriminative ability, satisfactory calibration, and robust accuracy indices preserved, we strongly advocate for the utilization of the MM-BM DDx model in clinical practice. Its implementation can facilitate the appropriate and timely referral of patients with a high likelihood of MM to hematologists, potentially reducing the time taken for MM diagnosis and treatment and ultimately leading to improved prognosis and survival.

## Data Sharing Statement

Supplementary information accompanies this paper can be found in Supplementary Materials.

## Ethical Approval

The ethical approval was exempted by the Ethical Committee from all contributing centers including six university-affiliated hospitals and medical schools participated in this study owing to the retrospective nature of deidentified data collection: Faculty of Medicine, Chiang Mai University, Khon Kaen University, Siriraj Hospital at Mahidol University, Queen Savang Vadhana Memorial Hospital, Khon Kaen Hospital, and Chulabhorn Hospital. All collected data were kept confidential and accessible to the investigators.

## Acknowledgments

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Weber KL. Evaluation of the adult patient (aged >40 years) with a destructive bone lesion. *J Am Acad Orthop Surg*. 2010;18(3):169–179. doi:10.5435/00124635-201003000-00006
2. Utzschneider S, Schmidt H, Weber P, Schmidt GP, Jansson V, Dürr HR. Surgical therapy of skeletal complications in multiple myeloma. *Int Orthop*. 2011;35(8):1209–1213. doi:10.1007/s00264-010-1127-0
3. Baykara M, Yildirim M. Differentiation of multiple myeloma and metastases with apparent diffusion coefficient map histogram analysis. *North Clin Istanb*. 2022;9(3):256–260.

4. Lee YJ, Jee WH, Ha KY, et al. MR distinction between multiple myeloma and metastasis involving the spine. *J Korean Radiol Soc*. 2016;44 (2):229–235. doi:10.3348/jkrs.2001.44.2.229

5. Kim HJ, Ryu KN, Choi WS, Choi BK, Choi JM, Yoon Y. Spinal involvement of hematopoietic malignancies and metastasis: differentiation using MR imaging. *Clin Imaging*. 1999;23(2):125–133. doi:10.1016/S0899-7071(99)00105-9

6. Kariyawasan CC, Hughes DA, Jayatillake MM, Mehta AB. Multiple myeloma: causes and consequences of delay in diagnosis. *QJM*. 2007;100 (10):635–640. doi:10.1093/qjmed/hcm077

7. Koshiaris C. Methods for reducing delays in the diagnosis of multiple myeloma. *Int J Hematol Oncol*. 2019;8(1):IJH13. doi:10.2217/ijh-2018-0014

8. Ryan C, Stoltzfus KC, Horn S, et al. Epidemiology of bone metastases. *Bone*. 2022;158:115783. doi:10.1016/j.bone.2020.115783

9. Zhou L, Yu Q, Wei G, et al. Measuring the global, regional, and national burden of multiple myeloma from 1990 to 2019. *BMC Cancer*. 2021;21 (1):606. doi:10.1186/s12885-021-08280-y

10. Phinyo P, Maihom T, Phanphaisarn A, et al. Development of a clinical diagnostic tool to differentiate multiple myeloma from bone metastasis in patients with destructive bone lesions (MM-BM DDx). *BMC Fam Pract*. 2020;21(1):215.

11. Phinyo P, Maihom T, Phanphaisarn A, et al. Correction: development of a clinical diagnostic tool to differentiate multiple myeloma from bone metastasis in patients with destructive bone lesions (MM-BM DDx). *BMC Prim Care*. 2022;23(1):189. doi:10.1186/s12875-022-01785-w

12. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279–289. doi:10.1016/j.jclinepi.2014.06.018

13. World Medical Association. World Medical Association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191–2194. doi:10.1001/jama.2013.281053

14. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med*. 2015;13(1):1. doi:10.1186/s12916-014-0241-z

15. 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):e538–548. doi:10.1016/S1470-2045(14)70442-5

16. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med*. 2021;40(19):4230–4251. doi:10.1002/sim.9025

17. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Med Res Methodol*. 2014;14(1):75. doi:10.1186/1471-2288-14-75

18. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol*. 2009;9:57. doi:10.1186/1471-2288-9-57

19. Greipp PR, San Miguel J, Durie BGM, et al. International staging system for multiple myeloma. *J Clin Oncol*. 2005;23(15):3412–3420. doi:10.1200/JCO.2005.04.242

20. Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol*. 2010;172(8):971–980. doi:10.1093/aje/kwq223

21. Xiong X, Wang J, Hu S, Dai Y, Zhang Y, Hu C. Differentiating between multiple myeloma and metastasis subtypes of lumbar vertebra lesions using machine learning–based radiomics. *Front Oncol*. 2021;11:1.

22. Park GE, Jee WH, Lee SY, et al. Differentiation of multiple myeloma and metastases: use of axial diffusion-weighted MR imaging in addition to standard MR imaging at 3T. *PLoS One*. 2018;13(12):e0208860. doi:10.1371/journal.pone.0208860

23. McGee S. Simplifying likelihood ratios. *J Gen Intern Med*. 2002;17(8):647–650.

24. Eleutherakis-Papaiakovou V, Bamias A, Gika D, et al. Renal failure in multiple myeloma: incidence, correlations, and prognostic significance. *Leuk Lymphoma*. 2007;48(2):337–341. doi:10.1080/10428190601126602

25. Clark AD, Shetty A, Soutar R. Renal failure and multiple myeloma: pathogenesis and treatment of renal failure and management of underlying myeloma. *Blood Rev*. 1999;13(2):79–90. doi:10.1016/S0268-960X(99)90014-0

26. Park S, Han B, Kim K, et al. Renal insufficiency in newly-diagnosed multiple myeloma: analysis according to International Myeloma Working Group consensus statement. *Anticancer Res*. 2014;34(8):4299–4306.

27. Cai Y, Zhao Y, Dai Q, Xu M, Xu X, Xia W. Prognostic value of the albumin–globulin ratio and albumin–globulin score in patients with multiple myeloma. *J Int Med Res*. 2021;49(3):0300060521997736. doi:10.1177/0300060521997736

28. Laudin GE, Levay PF, Coetzer B. Globulin fraction and albumin: globulin ratio as a predictor of mortality in a South African multiple myeloma cohort. *Int J Hematol Oncol*. 2020;9(3):IJH27. doi:10.2217/ijh-2020-0003

29. Thio QCBS, Karhade AV, Notman E, et al. Serum alkaline phosphatase is a prognostic marker in bone metastatic disease of the extremity. *J Orthop*. 2020;22:346–351. doi:10.1016/j.jor.2020.08.008

30. Karhade AV, Thio QCBS, Kuverji M, Ogink PT, Ferrone ML, Schwab JH. Prognostic value of serum alkaline phosphatase in spinal metastatic disease. *Br J Cancer*. 2019;120(6):640–646. doi:10.1038/s41416-019-0407-8

31. Koshiaris C, Bruel den AV, Nicholson BD, Lay-Flurrie S, Hobbs FR, Oke JL. Clinical prediction tools to identify patients at highest risk of myeloma in primary care: a retrospective open cohort study. *Br J Gen Pract*. 2021;71(706):e347–55. doi:10.3399/BJGP.2020.0697

32. Smith L, Carmichael J, Cook G, Shinkins B, Neal RD. Development and internal validation of a risk prediction model to identify myeloma based on routine blood tests: a case-control study. *Cancers*. 2023;15(3):975.

33. Annibali O, Petrucci MT, Santini D, et al. Alkaline phosphatase (alp) levels in multiple myeloma and solid cancers with bone lesions: is there any difference? *J Bone Oncol*. 2020;26:100338. doi:10.1016/j.jbo.2020.100338

34. Altman RL, Lin CT, Earnest M. Problem-oriented documentation: design and widespread adoption of a novel toolkit in a commercial electronic health record. *JAMIA Open*. 2023;6(1):ooad005. doi:10.1093/jamiaopen/ooad005

35. Jehn CF, Diel IJ, Overkamp F, et al. Management of metastatic bone disease algorithms for diagnostics and treatment. *Anticancer Res*. 2016;36(6):2631–2637.

36. Seesaghur A, Petruski-Ivleva N, Banks VL, et al. Clinical features and diagnosis of multiple myeloma: a population-based cohort study in primary care. *BMJ Open*. 2021;11(10):e052759. doi:10.1136/bmjopen-2021-052759

37. Mateos MV, Fink L, Koneswaran N, et al. Bone complications in patients with multiple myeloma in five European countries: a retrospective patient chart review. *BMC Cancer*. 2020;20(1):170. doi:10.1186/s12885-020-6596-y