Received: 25 April 2017 | Revised: 5 July 2017 | Accepted: 16 August 2017

DOI: 10.1002/ajh.24891

RESEARCH ARTICLE

WILEY AJH

# Comprehensive evaluation of the revised international staging system in multiple myeloma patients treated with novel agents as a primary therapy

Hyungwoo Cho[1] | Dok Hyun Yoon[2] | Jung Bok Lee[3] | Sung-Yong Kim[4] | Joon Ho Moon[5] | Young Rok Do[6] | Jae Hoon Lee[7] | Yong Park[8] | Ho Sup Lee[9] | Hyeon Seok Eom[10] | Ho-Jin Shin[11] | Chang-Ki Min[12] | Jin Seok Kim[13] | Jae-Cheol Jo[14] | Hye Jin Kang[15] | Yeung-Chul Mun[16] | Won Sik Lee[17] | Je-Jung Lee[18] | Cheolwon Suh[2] | Kihyun Kim[19] | and the Korean Multiple Myeloma Working Party

[1]Department of Internal Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; [2]Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; [3]Department of Clinical Epidemiology and Biostatistics, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; [4]Division of Hematology–Oncology, Department of Internal Medicine, Konkuk University School of Medicine, Konkuk University, Seoul, South Korea; [5]Department of Hematology–Oncology, Kyungpook National University Hospital, Daegu, South Korea; [6]Department of Hemato-Oncology, Keimyung University Dongsan Medical Center, Daegu, South Korea; [7]Division of Hematology/Oncology, Gachon University Gil Medical Center, Incheon, South Korea; [8]Division of Hematology–oncology, Department of Internal Medicine, Korea University School of Medicine, Seoul, South Korea; [9]Department of Internal Medicine, Kosin University College of Medicine, Busan, South Korea; [10]Department of Hematology–Oncology, Center for Hematologic Malignancy, National Cancer Center, Goyang-si, South Korea; [11]Department of Internal Medicine, Pusan National University Hospital, Busan, South Korea; [12]Department of Internal Medicine, Seoul St. Mary's Hospital, The Catholic University of Korea, Seoul, South Korea; [13]Department of Internal Medicine, Yonsei University College of Medicine, Severance Hospital, Seoul, South Korea; [14]Department of Hematology and Oncology, Ulsan University Hospital, University of Ulsan College of Medicine, Ulsan, South Korea; [15]Department of Internal Medicine, Korea Cancer Center Hospital, Korea Institute of Radiological and Medical Sciences, Seoul, South Korea; [16]Department of Internal Medicine, Ewha Womans University School of Medicine, Seoul, South Korea; [17]Department of Internal Medicine, Inje University Busan Paik Hospital, Busan, South Korea; [18]Department of Hematology-Oncology, Chonnam National University Hwasun Hospital, Hwasun, South Korea; [19]Division of Hematology/Oncology, Department of Medicine, Sungkyunkwan University School of Medicine, Samsung Medical Center, Seoul, South Korea

**Correspondence**
Kihyun Kim, MD, PhD, Division of Hematology/Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 50 Irwon-dong Gangnam-gu, Seoul, 135-710, Korea.
Email: kihyunkimk@gmail.com

## Abstract

The revised International Staging System (R-ISS) has recently been developed to improve the risk stratification of multiple myeloma (MM) patients over the ISS. We assessed the R-ISS in MM patients who were treated with novel agents as a primary therapy and evaluated its discriminative power and ability to reclassify patients from the ISS. A total of 514 newly diagnosed MM patients treated with novel agents including thalidomide, bortezomib, and lenalidomide as a primary therapy were included in this retrospective analysis. With a median follow-up duration of 42.3 months (range, 40.5–44.1), the median overall survival (OS) was 61.0 months. There was a significant difference in median OS (not reached, 60.9, and 50.1 months for stages 1, 2, and 3, respectively, $P < 0.001$) among the three stages of R-ISS. The C-statistic was significantly greater for R-ISS than for ISS (0.769 vs. 0.696, $P < 0.001$). The event NRI was −0.08 (95% confidence interval [CI], −0.18–0.01) and the non-event NRI was 0.05 (95% CI, −0.03–0.10), resulting in a total NRI of −0.03 (95% CI, −0.14–0.08, $P = 0.602$). The R-ISS performs well and has significantly better discriminative power than the ISS in MM patients treated with novel agents as a primary therapy. However, it does not better reclassify patients from the ISS, suggesting that there is still room to

Hyungwoo Cho and Dok Hyun Yoon equally contributed to this work and should be considered co-first authors.

improve the staging system. Moreover, new statistical measures for assessing and quantifying the risk prediction of new prognostic models are necessary in future studies.

## 1 | INTRODUCTION

Multiple myeloma (MM) is a plasma cell neoplasm characterized by the accumulation of clonal plasma cells. It accounts for 1% of all neoplastic diseases and ~13% of all hematologic malignancies.[1] The introduction of novel agents has significantly improved the prognosis of MM patients, but MM is still considered an incurable disease, and there is wide variation in patient survival because of its heterogeneity. Thus, the use of an accurate prognostic model is important for applying risk-adapted therapeutic strategies, predicting disease outcomes, and informing patients.

The International Staging System (ISS), which is based on serum beta 2-microglobulin (Sβ2M) and serum albumin,[2] has been the standard prognostic model for the past 10 years. However, the patient data used to develop the ISS were collected from 1981 through 2002. Thus, the majority of patients whose data were used have never been exposed to the novel agents. Moreover, the ISS does not directly incorporate the role of intrinsic myeloma cell variability at the molecular level, such as cytogenetic abnormalities.

Recently, the Revised ISS (R-ISS), which incorporates serum LDH and high-risk cytogenetic abnormality (CA) (*t*[4;14], *t*[14;16], and del [17p]) detected by interphase fluorescence in situ hybridization (iFISH) into ISS, was developed by the International Myeloma Working Group (IMWG) to overcome this limitation.[3] The R-ISS improved the stratification of patients into more homogeneous risk groups, and is now widely accepted as the new standard prognostic model for MM patients.

In this study, we assessed the R-ISS in Korean MM patients who were treated with novel agents (thalidomide, bortezomib, or lenalidomide) as a primary therapy and evaluated its discriminative power and ability to reclassify patients from the ISS classifications.

## 2 | METHODS

### 2.1 | Patients

Clinical and laboratory data were collected from newly diagnosed MM patients between January 2010 and August 2013 from 17 hospitals in Korea. The inclusion criteria were as follows: diagnosis of MM based on IMWG criteria,[4] information for the determination of ISS and R-ISS stage at diagnosis, and treatment with novel chemotherapeutic agents including thalidomide, bortezomib, and lenalidomide as part of a first-line chemotherapy protocol. Patients without complete LDH and iFISH results but with available ISS and R-ISS stages were included in the study (e.g., complete information on LDH and FISH analysis results were not required for ISS stage 2 patients, because these patients were reclassified as R-ISS stage 2 regardless of their LDH and iFISH results). Patients with asymptomatic (smoldering) MM, immunoglobulin M-related disorders, or primary amyloidosis were excluded.

Clinical data were obtained by reviewing medical records, and this study was approved by the institutional review boards of all participating institutions. Baseline data collected included age, sex, ISS stage, R-ISS stage, performance status (PS) according to the Eastern Cooperative Oncology Group scale, serum albumin, serum LDH, Sβ2M, serum creatinine, estimated glomerular filtration rate calculated using the Modification of Diet in Renal Disease Study formula, serum calcium, hemoglobin level, serum free light chain ratio, clonal bone-marrow plasma-cell percentage, and CA detected by iFISH. Serum levels of LDH were classified as normal or high per the laboratory definition of the normal range of each institution. The iFISH studies were performed on sorted or immunologically recognized plasma cells according to the iFISH methods of each institution. The presence of del(17p), *t*(4;14), or *t*(14;16) detected by iFISH was considered to indicate high-risk CA.

### 2.2 | Statistical analysis

Overall survival (OS) was calculated from the start date of chemotherapy until death from any cause or was censored at last follow-up. Progression-free survival (PFS) was defined as the time from the start date of chemotherapy until progression or death from any cause. Survival rates and corresponding standard errors were estimated using the Kaplan–Meier method, and survival curves were compared using the log-rank test. We evaluated the discriminatory power of the R-ISS vs. the ISS in terms of OS using the C-statistic,[5] which is also known as the area under the receiver operating characteristic curve, where greater values indicate greater discriminative power and a value of 1 indicates perfect discrimination. We report the C-statistic value at 3 years of follow-up, and histogram showing the C-statistic values of ISS and R-ISS at different periods of follow-up time was plotted. The concordance probability estimate was used to measure the discriminatory power of LDH and high-risk CA separately with respect to OS.[6] The integrated Brier score (IBS) was used to assess prediction error, and a lower IBS indicates a higher prediction accuracy.[7] Calibration for agreement was assessed by using D'Agostino-Nam test.[8]

The risk category reclassification by the R-ISS in terms of OS was assessed using the net reclassification improvement (NRI) metric.[9,10] Assuming independence between event and non-event individuals and following McNemar's[11] logic for significance testing in correlated proportions, we used a simple asymptotic test for the null hypothesis of NRI = 0. We report the NRI at 3 years of follow-up, separately for those with or without events, as well as the total NRI. A total NRI value of 2 indicates that all of the patients were correctly reclassified, whereas −2 indicates that all of the patients were incorrectly reclassified into another risk group.

## 3 | RESULTS

### 3.1 | Patient characteristics and treatments

A total of 568 newly diagnosed MM patients were treated with novel chemotherapeutic agents. Of these, 54 patients were excluded due to

insufficient data for ISS and R-ISS, and 514 patients formed the basis of this analysis. The baseline characteristics were similar between those with or without sufficient data for ISS and R-ISS, but patients with sufficient data less often received autologous stem cell transplantation (ASCT) (40.9% vs. 57.4%, $P = 0.020$), while the frequency of primary therapy with thalidomide-based (55.4% vs. 68.5%), bortezomib-based (42.0% vs. 29.6%) and lenalidomide-based (2.5% vs. 1.9%) regimen were not significantly different.

The baseline characteristics of the patients are presented in Table 1. The median age was 63 years (range, 32–86 years); 205 patients (39.9%) were >65 years of age. Although data for LDH and high-risk CAs were missing in 5 (1%) and 74 (14.4%) patients, respectively, ISS and R-ISS stages were available for all of the patients. In all, 177 patients (33.4%) had abnormal LDH levels, and 92 patients (17.9%) had high-risk CAs. All patients received novel agents as primary therapy, 298 (57.9%) received immunomodulatory drugs (55.4% thalidomide-based, 2.4% lenalidomide-based), and 216 (42.0%) received bortezomib-based primary therapy. In all, 210 patients (40.9%) received ASCT. With a median follow-up duration of 42.3 months (range, 40.5–44.1 months), median PFS was 25.0 months and median OS was 61.0 months. There was no significant difference in median OS between patients with or without sufficient data for ISS and R-ISS (median OS 61.0 months vs. 49.0 months, $P = 0.850$).

### 3.2 | ISS distribution and survival outcome

One hundred thirteen patients (22.0%) were ISS stage 1, 210 (40.9%) were ISS stage 2, and 191 (37.2%) were ISS stage 3. The median PFS according to ISS stages 1, 2, and 3 were 33.8, 25.4, and 18.0 months, respectively (Figure 1a). There was no significant difference between the median PFS of ISS stage 1 and stage 2 patients (stage 1 vs. stage 2, $P = 0.067$; stage 1 vs. stage 3, $P < 0.001$; stage 2 vs. stage 3, $P = 0.019$). Median OS according to ISS stage were not reached for ISS stages 1 and 2 and was 48.0 months for ISS stage 3 (Figure 1b). There was no significant difference between the median OS of ISS stage 1 and stage 2 patients (stage 1 vs. stage 2, $P = 0.334$; stage 1 vs. stage 3, $P = 0.001$; stage 2 vs. stage 3, $P = 0.003$).

### 3.3 | R-ISS distribution and survival outcome

Per the R-ISS, 64 patients (12.5%) were rated as R-ISS stage 1, 338 (65.8%) were R-ISS stage 2, and 112 (21.8%) were R-ISS stage 3. Among the 113 patients with ISS stage 1, 49 patients (43.4%) were reclassified as R-ISS stage 2, and among 191 patients with ISS stage 3, 79 patients (41.4%) were reclassified as R-ISS stage 2. Median PFSs according to R-ISS stages 1, 2, and 3 were 40.1, 23.9, and 18.3 months, respectively (Figure 2a). There were significant differences in median PFS among the three stages of the R-ISS (stage 1 vs. stage 2, $P = 0.014$; stage 1 vs. stage 3, $P = 0.001$; stage 2 vs. stage 3, $P = 0.045$). The median OS was not reached for R-ISS stage 1, was 60.9 months for R-ISS stage 2, and was 50.1 months for R-ISS stage 3 (Figure 2b). There were significant differences in median OS between the three stages of the R-ISS (stage 1 vs. stage 2, $P = 0.002$; stage 1 vs. stage 3, $P < 0.001$; stage 2 vs. stage 3, $P = 0.044$).

### 3.4 | Comparison of the performance of ISS and R-ISS

As both the ISS and the R-ISS were primarily developed to stratify patients into different risk groups in terms of OS, C-statistics, Brier score, D'Agostino-Nam test and NRI analysis were used to assess discriminative power, prediction accuracy, calibration for agreement and risk category reclassification, respectively, in terms of OS but not PFS. The value of the C-statistic was significantly greater for the R-ISS than the ISS (0.769 vs. 0.696, $P < 0.001$) (Supporting Information Figure S1). The IBS was 0.190 and 0.191 for ISS and R-ISS, respectively. R-ISS showed a good calibration with a D'Agostino-Nam test ($\chi^2 = 2.222$, $P = 0.334$), whereas ISS showed a poor calibration with a D'Agostino-Nam test ($\chi^2 = 11.443$, $P = 0.004$). The results of NRI analyses are presented in Table 2. Among patients with ISS stage 1, the event (death) rate of patients who were reclassified as R-ISS stage 2 was 41.3%, whereas that of patients reclassified as R-ISS stage 1 was 15.1%. The event (death) rate of patients with ISS stage 2, of whom all were reclassified as R-ISS stage 2, was 28.6%. Among patients with ISS stage 3, the event (death) rate of patients who were reclassified as R-ISS stage 2 was 42.3%, and that of patients reclassified as R-ISS stage 3 was 43.5%. The event NRI was −0.08 (95% confidence interval [CI], −0.18–0.01) and the non-event NRI was 0.05 (95% CI, −0.03–0.10), resulting in a total NRI of −0.03 (95% CI, −0.14–0.08, $P = 0.602$).

### 3.5 | Impact of abnormal LDH and high-risk CA on survival outcome

The concordance probability estimate was 0.543 (Standard error [SE], 0.012) for high-risk CA and 0.549 (SE, 0.016) for abnormal LDH, indicating that the discriminatory power was acceptable for both high-risk CA and abnormal LDH. The impact of abnormal LDH and high-risk CA on OS was evaluated within each ISS stage. Of patients with ISS stage 2, 41 who were lacking complete data for LDH and iFISH results were excluded from this subgroup analysis. In ISS stage 1, although the median OS was not reached for patients with either high-risk factor (n = 49) and patients with no high-risk factor ($n = 64$), the 3-year OS rates were 58.2% vs. 84.5%, respectively ($P = 0.002$) (Supporting Information Figure S2a). In ISS stage 2, there were no significant differences in OS between patients with either high-risk factor ($n = 67$) and patients with no high-risk factor ($n = 102$), with a median OS of 60.0 months vs. 61.3 months and a 3-year OS rate of 65.9% vs. 73.3%, respectively ($P = 0.346$) (Supporting Information Figure S2b). Similarly, in ISS stage 3, there were no significant differences in OS between patients with either high-risk factor ($n = 112$) and patients with no high-risk factor ($n = 79$), with a median OS of 50.1 months vs. 46.6 months and a 3-year OS rate of 56.2% vs. 57.3%, respectively ($P = 0.838$) (Supporting Information Figure S2c).

The proportion of high LDH (33.4%) in the current study was high compared to the IMWG cohort, in which only 13% of the patients had high LDH. Because LDH could not discriminate survival outcome of patients with ISS stage 3, we additionally analyzed OS of ISS stage 3 patients using higher cutoff for LDH. There was no significant difference in median OS of patients with LDH > 1.5×[upper normal limit]

**TABLE 1** Characteristics of patients

| Characteristics | No. of patients (%) |
|---|---|
| Age, median (range) | 63 (32–86) |
| Age | |
| ≤65 | 309 (60.1%) |
| >65 | 205 (39.9%) |
| Missing | 0 (0.0%) |
| Sex | |
| Male | 281 (54.7%) |
| Female | 233 (45.3%) |
| Missing | 0 (0.0%) |
| Lactate dehydrogenase | |
| Normal | 332 (64.6%) |
| High | 177 (34.4%) |
| Missing | 5 (1.0%) |
| Serum albumin | |
| <3.5 g $dL^{-1}$ | 252 (49.0%) |
| ≥3.5 g $dL^{-1}$ | 262 (51.0%) |
| Missing | 0 (0.0%) |
| Serum beta 2-microglobulin | |
| <3.5 μg $mg^{-1}$ | 196 (38.1%) |
| ≥3.5 μg $mg^{-1}$ | 318 (61.9%) |
| Missing | 0 (0.0%) |
| Serum creatinine | |
| <2.0 mg $dL^{-1}$ | 417 (81.1%) |
| ≥2.0 mg $dL^{-1}$ | 97 (18.9%) |
| Missing | 0 (0.0%) |
| eGFR, median (range in mL $min^{-1}$ 1.73 $m^{-2}$) | 66.5 (3–244) |
| eGFR | |
| <30 mL $min^{-1}$ 1.73 $m^{-2}$ | 162 (31.5%) |
| ≥30 mL $min^{-1}$ 1.73 $m^{-2}$ | 327 (63.6%) |
| Missing | 25 (4.9%) |
| ECOG PS ≥2 | |
| 0–1 | 393 (76.5%) |
| ≥2 | 120 (23.3%) |
| Missing | 1 (0.2%) |
| Hemoglobin | |
| <10 g $dL^{-1}$ | 287 (55.8%) |
| ≥10 g $dL^{-1}$ | 223 (43.4%) |
| Missing | 4 (0.8%) |
| Calcium | |
| ≤11 mg $dL^{-1}$ | 458 (89.1%) |
| >11 mg $dL^{-1}$ | 52 (10.1%) |
| Missing | 4 (0.8%) |
| BMPC | |
| <60% | 360 (70.0%) |
| ≥60% | 141 (27.4%) |
| Missing | 13 (2.5%) |
| sFLC ratio | |
| 0.01–100 | 282 (54.9%) |
| <0.01 or >100 | 214 (41.6%) |
| Missing | 18 (3.5%) |
| ASCT | |
| No | 303 (58.9%) |
| Yes | 210 (40.9%) |
| Missing | 1 (0.2%) |
| High risk cytogenetics[a] detected by iFISH | |
| No | 348 (67.7%) |
| Yes | 92 (17.9%) |
| Missing | 74 (14.4%) |
| Del(17) detected by iFISH | |
| No | 420 (81.7%) |
| Yes | 27 (5.3%) |
| Missing | 67 (13.0%) |
| *t*(4;14) detected by iFISH | |
| No | 392 (76.3%) |
| Yes | 69 (13.4%) |
| Missing | 53 (10.3%) |
| *t*(14;16) detected by iFISH | |
| No | 405 (78.8%) |
| Yes | 34 (6.6%) |
| Missing | 75 (14.6%) |
| Chemotherapy | |
| Thalidomide-based | 285 (55.4%) |
| Lenalidomide-based | 216 (42.0%) |
| Bortezomib-based | 13 (2.5%) |

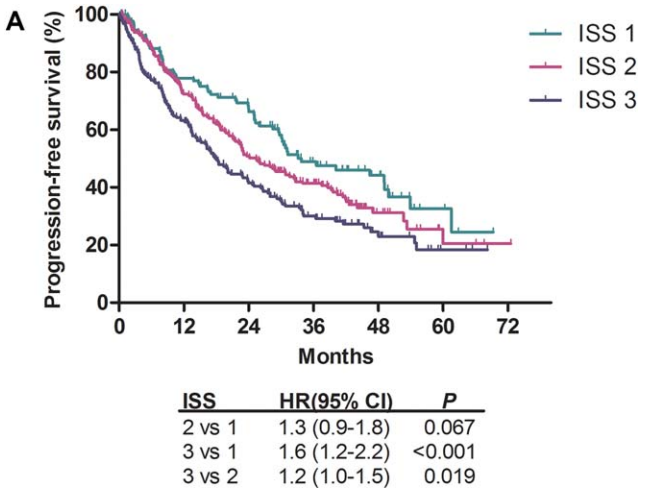
Abbreviations: ECOG PS = Eastern Cooperative Oncology Group performance status, BMPC = clonal bone-marrow plasma-cell percentage, sFLC = serum free light chain, iFISH = interphase fluorescence in situ hybridization.
[a]del17p, *t*(4;14) or *t*(14;16).

and with LDH ≤1.5×[upper normal limit] (median OS 48.0 [95% CI, 36.1–59.9] vs. 46.8 [95% CI, 14.6–78.9], $P = 0.405$).

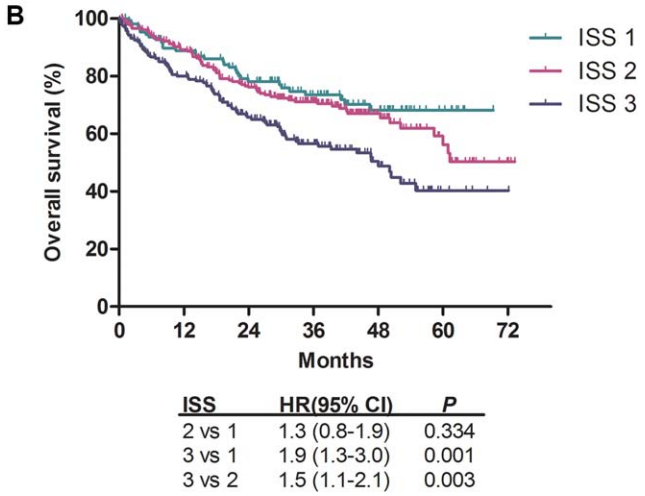
## 4 | DISCUSSION

A robust prognostic model is necessary to categorize patients into homogeneous risk groups, particularly in diseases with heterogeneous outcomes such as MM. Moreover, the validation of a prognostic model is an important step for its acceptance as an effective prognostic tool and use in daily clinical practice. The current study demonstrates that the R-ISS is well validated in Korean patients with newly diagnosed MM who were treated with novel agents as a primary therapy.

Using the R-ISS, 12.5% ($n = 64$) of our cases were classified as R-ISS stage 1, 65.8% ($n = 338$) as R-ISS stage 2, and 21.8% ($n = 112$) as R-ISS stage 3, and there were significant differences in PFS and OS rates between the groups. This is in line with the results of recent studies, which have also demonstrated that the R-ISS was well validated in MM patients in the era of novel agents.[12,13] The ISS, the former standard prognostic model, was also prognostic among our MM population in terms of both PFS ($P = 0.001$) and OS ($P = 0.001$). The IBS demonstrated that the predictive accuracy was acceptable for both ISS and R-ISS. However, the prognostic significance of ISS is mainly due to the poorer survival outcomes of ISS stage 3 patients. There were no significant differences in the PFS and OS rates between ISS stage 1 and 2 patients ($P = 0.067$ for PFS and $P = 0.334$ for OS, respectively).




| ISS | HR(95% CI) | P |
|---|---|---|
| 2 vs 1 | 1.3 (0.9-1.8) | 0.067 |
| 3 vs 1 | 1.6 (1.2-2.2) | <0.001 |
| 3 vs 2 | 1.2 (1.0-1.5) | 0.019 |




| ISS | HR(95% CI) | P |
|---|---|---|
| 2 vs 1 | 1.3 (0.8-1.9) | 0.334 |
| 3 vs 1 | 1.9 (1.3-3.0) | 0.001 |
| 3 vs 2 | 1.5 (1.1-2.1) | 0.003 |

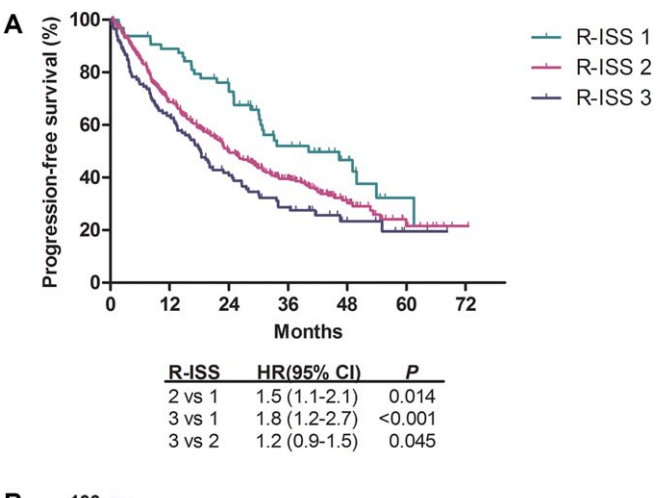
**FIGURE 1** Survival outcomes according to ISS. (a) Progression-free survival. (b) Overall survival [Color figure can be viewed at wileyonlinelibrary.com]

Moreover, the R-ISS demonstrated a significant improvement of discriminative power compared to the ISS in the C-statistic, although the magnitude of improvement was small (0.769 vs. 0.696, $P < 0.001$). In addition, R-ISS showed a better calibration performance compared to ISS ([$\chi^2 = 2.222$, $P = 0.334$] vs. [$\chi^2 = 11.443$, $P = 0.004$]).

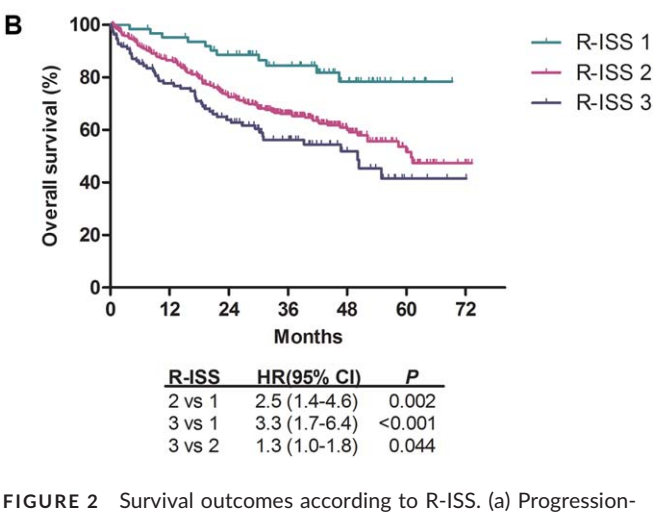
In contrast to the results of the log-rank test and C-statistic analyses, NRI analyses indicated that the R-ISS did not show better prediction of survival outcomes than the ISS. The event rate of ISS stage 1 patients who were reclassified as R-ISS stage 2 was higher than that of patients reclassified as R-ISS stage 1 (41.3% vs. 15.1%), suggesting that higher-risk ISS stage 1 patients were appropriately reclassified as R-ISS stage 2. However, the event rate of ISS stage 3 patients who were reclassified as R-ISS stage 2 was similar to that of patients reclassified as R-ISS stage 3 (42.4% vs. 43.5%), suggesting that the R-ISS inappropriately reclassified ISS stage 3 patients. This resulted in an overall NRI of −0.03 (95% CI, −0.14–0.08, $P = 0.602$), meaning that the R-ISS did not better reclassify patients from the ISS. Another reason for this unexpected result might be that none of the ISS stage 2 patients were reclassified into another risk group, which resulted in both event and nonevent NRI values being close to zero; consequently, total NRI value was close to zero. Moreover, this led to an imbalance in the number of patients among the three stages of R-ISS, with more than half of the patients classified as R-ISS stage 2 ($n = 338$, 65.8%).

The subgroup analysis of each ISS risk group had similar results as the NRI analysis. In ISS stage 1 patients, there was a significant difference in OS between patients with either high-risk factor and those with no high-risk factor. However, in ISS stage 2 or 3 patients, there were no significant differences in OS between patients with either high-risk factor and no high-risk factor. Moreover, there were no significant differences in OS between ISS stage 1 patients with either high-risk factor and ISS stage 3 patients with either high-risk factor (3-year OS rate 58.2% vs. 56.2%, $P = 0.598$) (Supporting Information Figure S3). This suggests that the presence of either abnormal LDH or high-risk cytogenetics resulted in poor OS in ISS stage 1 patients but not in ISS stage 2 or 3 patients.

Our current study results demonstrate that although the R-ISS is well validated and has significantly better discriminative power than the ISS, there remain several limitations with this system that suggest there is still room for improvement. To ensure uniform availability, only three widely available cytogenetic markers, del(17p), *t*(4;14),




| R-ISS | HR(95% CI) | P |
|---|---|---|
| 2 vs 1 | 1.5 (1.1-2.1) | 0.014 |
| 3 vs 1 | 1.8 (1.2-2.7) | <0.001 |
| 3 vs 2 | 1.2 (0.9-1.5) | 0.045 |




| R-ISS | HR(95% CI) | P |
|---|---|---|
| 2 vs 1 | 2.5 (1.4-4.6) | 0.002 |
| 3 vs 1 | 3.3 (1.7-6.4) | <0.001 |
| 3 vs 2 | 1.3 (1.0-1.8) | 0.044 |

**FIGURE 2** Survival outcomes according to R-ISS. (a) Progression-free survival. (b) Overall survival [Color figure can be viewed at wileyonlinelibrary.com]

**TABLE 2** Reclassification table

| | R-ISS ($n = 514$) | | | | % reclassified into new risk category | | |
|---|---|---|---|---|---|---|---|
| ISS | 1 | 2 | 3 | Total (% of ISS) | Lower | Higher | Total |
| 1 | 64 | 49 | – | 113 (22.0%) | 0.0% | 43.4% | 43.4% |
| Event rate at 36 months | 15.1% | 41.3% | – | | | | |
| 2 | – | 210 | – | 210 (40.9%) | 0.0% | 0.0% | 0.0% |
| Event rate at 36 months | – | 28.6% | – | | | | |
| 3 | – | 79 | 112 | 191 (37.2%) | 41.4% | 0.0% | 41.4% |
| Event rate at 36 months | – | 42.3% | 43.5% | | | | |
| Total (% of R-ISS) | 64 (12.5%) | 338 (65.8%) | 112 (21.8%) | | | | |

and *t*(14;16), were used in the R-ISS. However, MM patients with other CAs, such as gain 1q and *t*(14;20), are also known to have a poor prognosis, so these factors should also be included in the staging system.[14] More recently, gene expression profiling (GEP) has been extensively investigated as a potential tool for the assessment of risk in MM, and several GEP classifiers have been developed.[15–17] Studies have demonstrated that patients classified as being in a high-risk group based on these GEP classifiers have shorter survival than the low-risk group. Therefore, although GEP is quite complex and costly for routine use, it can provide additional prognostic value to the staging system.

Most previous studies on prognostic models for MM, including those that have investigated the ISS, the R-ISS, and GEP, have focused only on tumor-related prognostic factors.[2,3,15–17] However, patient-related factors such as age, PS, and renal function still play important roles as prognostic factors for MM patients.[18–20] Moreover, in diseases such as lymphoma, tumor-related and patient-related factors are combined to constitute prognostic models used in daily clinical practice.[21,22] Thus, a combinatorial approach using both tumor-related and patient-related prognostic factors should also be considered an important strategy for the improvement of the staging system.

Researchers continue to seek new risk factors that can predict the survival rates of certain diseases and try to incorporate them into risk-assessment algorithms to develop new prognostic models. However, the critical question arises of how new prognostic models are to be evaluated. In MM, most studies have focused on the validation of new prognostic models. To the best of our knowledge, our present study is the first to evaluate the discriminative power, prediction accuracy, calibration for agreement and ability of the R-ISS to reclassify patients from the ISS using various statistical measures. These results suggest that in addition to validation, other statistical measures are necessary to assess and quantify improvement in risk prediction offered by new prognostic models.

This study had several limitations of note. Only Korean patients were analyzed, and so our findings should be further investigated in a different population of MM patients. In addition, our median follow-up duration was relatively short. Furthermore, as expected for any retrospective study, there may have been a selection bias. Moreover, information on other prognostic factors such as 1q abnormalities, *t*(14;20), or GEP were not available. Finally, because our primary intent was to validate the R-ISS and to investigate its discriminative power and ability to reclassify patients from the ISS, patients with available ISS and R-ISS stages but without complete data for LDH and FISH results were also included in analyses. However, as only data for ISS and R-ISS stages were required for validation, C-statistics, and NRI analyses, it is unlikely that patients with incomplete data for LDH and FISH influenced our results.

In conclusion, the R-ISS performs well and has significantly better discriminative power than the ISS in MM patients treated with novel agents as a primary therapy. However, it does not better reclassify ISS stage 3 patients, suggesting that there is still room for its improvement. Thus, further studies to improve the staging system in MM patients are clearly warranted. Moreover, other statistical measures for assessing and quantifying the risk prediction of new prognostic models are necessary in future studies.

## ACKNOWLEDGMENTS

The authors thank the clinical research coordinators who contributed to this study.

## ORCID

*Hyungwoo Cho* http://orcid.org/0000-0003-4307-2795
*Jin Seok Kim* http://orcid.org/0000-0001-8986-8436

## REFERENCES

[1] Palumbo A, Anderson K. Multiple myeloma. *N Engl J Med.* 2011; 364:1046–1060.

[2] Greipp PR, San Miguel J, Durie BGM, et al. International staging system for multiple myeloma. *J Clin Oncol.* 2005;23(15):3412–3420.

[3] Palumbo A, Loiseau HA, Oliva S, et al. Revised international staging system for multiple myeloma: a report from International Myeloma Working Group. *JCO.* 2015;33:2863–2869.

[4] Kyle RA, Rajkumar SV. Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma. *Leukemia.* 2009;23:3–9.

[5] Pencina MJ, D' Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statist Med*. 2004;23(13):2109–2123.

[6] Gonen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*. 2005;92(4):965–970.

[7] Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biom J*. 2008;50(4):457–479.

[8] Demler OV, Paynter NP, Cook NR. Tests of calibration and goodness-of-fit in the survival setting. *Stat Med*. 2015;34(10):1659–1680.

[9] Pencina MJ, D' Agostino RB, Vasan RS, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statist Med*. 2007;27(2):157–172.

[10] Pencina MJ, D' 'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statist Med*. 2010;30(1):11–21.

[11] Mcnemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 1947;12(2):153–157.

[12] Kastritis E, Terpos E, Roussou M, et al. Evaluation of the revised international staging system (R-ISS) in an independent cohort of unselected patients with multiple myeloma. *Haematologica*. 2017;102:593–599.

[13] Zepeda VHJ, Duggan P, Neri P, et al. Revised international staging system applied to real world multiple myeloma patients. *Clin Lymph Myeloma Leuk*. 2016;16:511–518.

[14] Mikhael JR, Dingli D, Roy V, et al. Management of newly diagnosed symptomatic multiple myeloma: updated mayo stratification of myeloma and risk-adapted therapy (mSMART) consensus guidelines 2013. *Mayo Clin Proc*. 2013;88(4):360–376.

[15] Kuiper R, Broyl A, de Knegt Y, et al. A gene expression signature for high-risk multiple myeloma. *Leukemia*. 2012;26(11):2406–2413.

[16] Shaughnessy JD, Zhan F, Burington BE, et al. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*. 2007;109:2276–2284.

[17] Kuiper R, van Duin M, van Vliet MH, et al. Prediction of high-and low-risk multiple myeloma based on gene expression and the International Staging System. *Blood*. 2015;126:1996–2004.

[18] Kleber M, Ihorst G, Terhorst M, et al. Comorbidity as a prognostic variable in multiple myeloma: comparative evaluation of common comorbidity scores and use of a novel MM–comorbidity score. *Blood Cancer J*. 2011;1(9):e35.

[19] Kyle RA, Gertz MA, Witzig TE, et al. Review of 1027 Patients with newly diagnosed multiple myeloma. *Mayo Clin Proc*. 2003;78(1):21–33.

[20] Palumbo A, Bringhen S, Mateos MV, et al. Geriatric assessment predicts survival and toxicities in elderly myeloma patients: an International Myeloma Working Group report. *Blood*. 2015;125:2068–2074.

[21] Shipp MA, Harrington DP, Anderson JR, et al. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med*. 1993;329:987–994.

[22] Solal-Celigny P, Roy P, Colombat P, et al. Follicular lymphoma international prognostic index. *Blood*. 2004;104(5):1258–1265.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Cho H, Yoon DH, Lee JB, et al. Comprehensive evaluation of the revised international staging system in multiple myeloma patients treated with novel agents as a primary therapy. *Am J Hematol*. 2017;92:1280–1286. https://doi.org/10.1002/ajh.24891