PCR134

Psychometric Properties of the Multiple Myeloma Symptom and Impact Questionnaire in Patients With Relapsed/Refractory Multiple Myeloma: Analysis of Phase 2 CARTITUDE-2 Study Cohorts A, B, and C

María-Victoria Mateos[1], Adam D Cohen[2], Yaël C Cohen[3], Mounzer Agha[4], Jesús San-Miguel[5], Shambavi Richard[6], Niels WCJ van de Donk[7], André De Champlain[8], Eva G Katz[9], Charles Iaconangelo[8], Kevin C De Braganca[10], Jordan M Schecter[10], Helen Varsos[10], Christina Corsale[10], William Deraedt[11], Mythili Koneru[12], Octavio Costa Filho[12], Muhammad Akram[12], Katharine S Gries[9]

[1]Hospital Universitario de Salamanca, Instituto de Investigación Biomédica de Salamanca (IBSAL), Centro de Investigación del Cáncer (IBMCC-USAL,CSIC), Salamanca, Spain; [2]Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [3]Tel-Aviv Sourasky (Ichilov) Medical Center, and Faculty of Medical & Health Sciences, Tel Aviv University, Tel Aviv, Israel; [4]UPMC Hillman Cancer Center, Pittsburgh, PA, USA; [5]Cancer Center Clínica Universidad Navarra, Pamplona, Spain; [6]Icahn School of Medicine at Mount Sinai, New York, NY, USA; [7]Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, Netherlands; [8]Janssen Global Services, Horsham, PA, USA; [9]Janssen Global Services, Raritan, NJ, USA; [10]Janssen Research & Development, Raritan, NJ, USA; [11]Janssen Research & Development, Beerse, Belgium; [12]Legend Biotech USA Inc., Somerset, NJ, USA

Key Takeaway

The MySIm-Q symptom score is reliable, valid, and responsive to change based on preliminary analyses from CARTITUDE-2 Cohorts A, B, and C, supporting its use as a fit-for-purpose PRO instrument in MM studies

Conclusions

- MySIm-Q symptom score internal consistency (Cronbach's α-coefficient, 0.89) and test-retest reliability (ICC[2,1], 0.81), 2 indicators of reliability, showed evidence of reproducibility
- Concurrent validity supported that the MySIm-Q symptom score evaluated similar constructs to existing PRO measures, and known-groups validity showed that the MySIm-Q symptom score differentiated between disease severity states
- Anchor- and distribution-based MWPC analyses demonstrate that the MySIm-Q symptom score is responsive to change and is capable of discriminating between patients maintaining improvement or declining in their condition or health state
- More extensive analysis of MySIm-Q symptom and impact measurement properties using data from the phase 3 CARTITUDE-4 study are forthcoming

Acknowledgments

The authors, Janssen, and Legend Biotech thank the patients who participated in this study, the staff members at the study sites, the data and safety monitoring committee, and the staff members involved in data collection and analyses. This study was funded by Janssen Research & Development, LLC, and Legend Biotech USA Inc. Medical writing support was provided by Lauren D'Angelo, PhD, of Eloquent Scientific Solutions, and funded by Janssen Global Services, LLC.

Disclosures

M-VM reports honoraria from AbbVie/Genentech, Amgen, Celgene, GSK, Janssen-Cilag, Sanofi, and Takeda; and has served in a consulting/advisory role for AbbVie, Amgen, Celgene, GSK, Janssen-Cilag, Pfizer, Regeneron, Roche/Genentech, and Takeda. **ADC** reports a consulting role for AbbVie, Arcelix, BMS, GSK, Ichnos Sciences, Iteos Therapeutics, Janssen, Novartis, Pfizer, Roche/Genentech, and Takeda; has received research funding from Genentech/Roche, GSK, Janssen, and Novartis; reports travel, accommodations, and expenses from AbbVie, BMS, Ichnos Sciences, and Janssen; and has patent and royalty interests from Novartis. **YCC** served in a consulting or advisory role for Amgen, GSK, and Janssen; received honoraria from BMS, Janssen, and Roche; served on a speakers' bureau for AbbVie and GSK; received research funding (institution) from Takeda; and reports travel, accommodations, and expenses from Sanofi Aventis GmbH. **MA** has stock and other ownership interests in GenCART, Inc. **JS-M** has received honoraria and has served in a consulting/advisory role for AbbVie, Amgen, BMS, Celgene, GSK, Haemalogix, Karyopharm, MSD, Pfizer, Roche, Regeneron, Sanofi, SecuraBio, and Takeda; and has received honoraria/served in a consulting role for Janssen and Novartis. **SR** reports honoraria from BMS and Janssen; reports a consulting role for Genentech; and received research funding from C4 Therapeutics, Gracell Biotechnologies, Heidelberg Pharma, and Janssen. **NWCJvdD** has served in a consulting/advisory role for AbbVie, Adaptive Biotechnologies, Amgen, Bayer, BMS, Celgene, Janssen, Novartis, Pfizer, Roche, Servier, and Takeda; and has received research funding from Amgen, BMS, Celgene, Cellectis, Janssen, and Novartis. **ADC**, **EGK**, **CI**, **KCDB**, **JMS**, **HV**, **CC**, **WD**, and **KSG** are employees of Janssen. **MK**, **OCF**, and **MA** are employees of Legend Biotech USA, Inc.

Introduction

- Multiple myeloma (MM) is an incurable hematologic malignancy characterized by symptoms of bone pain, fatigue, and reduced physical and cognitive functioning that impact patient health-related quality of life (HRQoL)[1-3]
- Patient-reported outcome (PRO) instruments used in clinical practice were developed and validated prior to recent therapeutic advances (eg, chimeric antigen receptor [CAR] T-cell therapy); therefore, they may not reflect the disease experience of MM patient populations undergoing treatment with newer therapies
- The Multiple Myeloma Symptom and Impact Questionnaire (MySIm-Q) is a newly developed and validated MM-specific PRO instrument designed to account for the changing treatment landscape in MM[4]
  - MySIm-Q measures disease-related symptoms and impacts due to treatment modalities with differing mechanisms of action available in clinical practice today
  - It can be collected alongside other PRO instruments to measure and complement core PRO-based efficacy and tolerability endpoints in cancer clinical trials
- Here, we describe measurement properties of the MySIm-Q symptom score using data from the phase 2, multicohort, open-label, multicenter CARTITUDE-2 study

Methods

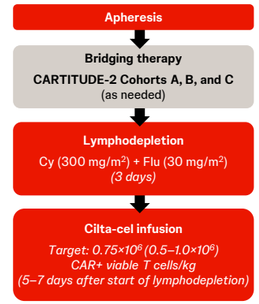
Key eligibility criteria and study design (Figure 1)

- Measurement properties of the MySIm-Q symptom score were assessed using data from Cohorts A, B, and C of CARTITUDE-2
  - **Cohort A**: comprised patients with lenalidomide-refractory MM and 1–3 prior lines of therapy (LOT), including a proteasome inhibitor (PI) and immunomodulatory drug (IMiD)[5]
  - **Cohort B**: comprised patients with 1 prior LOT, including a PI and IMiD, and disease progression ≤12 months after autologous stem cell transplant (ASCT) or frontline antimyeloma therapy for patients without ASCT[5]
  - **Cohort C**: comprised patients previously treated with a PI, IMiD, anti-CD38 antibody, and noncellular B-cell maturation antigen–directed therapy (eg, antibody-drug conjugate and bispecific T-cell engager)[6]

Assessments and statistical analysis

- **Internal consistency** was assessed at screening using Cronbach's α-coefficient; a threshold of >0.70 was selected as acceptable internal consistency[7]
- **Test-retest reliability** was assessed using a 2-way random intraclass correlation coefficient (ICC[2,1]),[8] with values ≥0.70 considered acceptable test-retest reliability

Figure 1: CARTITUDE-2 study design



Cy, cyclophosphamide; Flu, fludarabine.

- **Concurrent validity** (convergent validity) was considered supportive if Spearman correlations with the European Organisation for Research and Treatment of Cancer quality of life questionnaire core 30 (EORTC QLQ-C30) domain and items scores at screening were ≥0.4[9]
- **Known-groups validity** (discriminant validity) was determined at screening; Patient Global Impression of Severity (PGIS) and EORTC QLQ-C30 Global Health Status (GHS)/QoL domain items 29 and 30 were used as known-group validators
  - Differences calculated using a t-test, the mean differences and corresponding 95% confidence intervals (CIs), *P* values, and semi-partial ω^2 effect-size estimates
- **Meaningful within-patient change (MWPC)** estimates were calculated using an anchor-based approach and supplemented with an estimate derived from a distribution-based approach
  - Anchor: Patient Global Impression of Change (PGIC) at day 100
    - Median change scores were stratified by anchor groups of "minimal improvement," "no change," and "minimal deterioration"
    - Anchor analyses were further supported visually by empirical cumulative distribution functions (eCDFs) and empirical probability density functions (ePDFs)

Results

Study population

- A total of 82 patients completed MySIm-Q assessments (cohort A, n=43; cohort B, n=19; cohort C, n=20)

Measurement properties

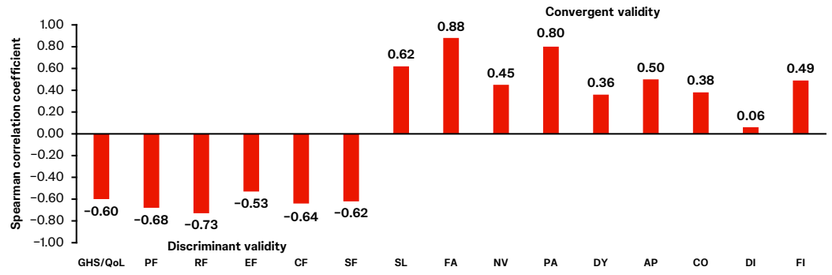
- Internal consistency was acceptable; Cronbach's α-coefficient was 0.89 for MySIm-Q symptom score
  - Item to MySIm-Q symptom score correlations ranged from 0.35–0.84 (**Table 1**)

Table 1: Internal consistency for the MySIm-Q symptom score

	Item to MySIm-Q symptom score correlation
Worst pain back	0.60
Worst pain leg	0.70
Worst pain other area	0.49
Worst numbness/tingling in hands and feet	0.35
Low energy	0.84
Tire easily	0.80
Muscle weakness	0.74
Trouble with sleep	0.55
Poor appetite	0.55
Difficulty with memory	0.50
Difficulty concentrating	0.62

- Test-retest reliability for the MySIm-Q symptom score was adequate; the ICC(2,1) was 0.81, exceeding the minimally acceptable value of 0.70
- MySIm-Q symptom score demonstrated acceptable concurrent validity with existing symptom and impact measures from the EORTC QLQ-C30 (**Figure 2**)
  - All correlations except the diarrhea domain met or exceeded the minimum correlation requirement of |r| ≥0.4, or were within rounding error of this criterion, with a high proportion exceeding | 0.60 |

Figure 2: Concurrent validity for the MySIm-Q symptom score



AP, appetite loss; CF, cognitive functioning; CO, constipation; DI, diarrhea; DY, dyspnea; EF, emotional functioning; FA, fatigue; FI, financial difficulties; NV, nausea and vomiting; PA, pain; PF, physical functioning; RF, role functioning; SF, social functioning; SL, insomnia.

- Known-groups validity for the MySIm-Q symptom score was supported
  - Discrimination across disease severity groups based on PGIS was shown for every category, with each PGIS category exhibiting incrementally larger differences from the "never" (reference) category (**Table 2**)
    - The "very severe" category was the exception; however, the sample size comprised 3 patients and estimates were unstable
- Discrimination evidence across disease severity groups based on EORTC QLQ-C30 GHS/QoL domain items were generally supportive

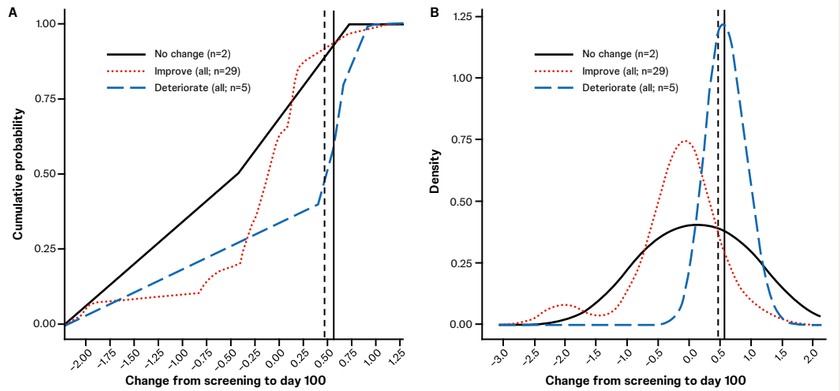
Table 2: Known-groups validity of MySIm-Q symptom score

Known-groups validator	Estimate (95% CI)	*P* value	ω^2
PGIS			0.946
Never vs mild	0.53 (0.23, 0.84)	0.0010	–
Never vs moderate	1.17 (0.83, 1.52)	<0.0001	–
Never vs severe	1.55 (1.12, 1.98)	<0.0001	–
Never vs very severe	1.41 (0.65, 2.18)	0.0005	–
EORTC GH29			0.921
Excellent vs very poor	1.34 (0.52, 2.17)	0.0018	–
Excellent vs GH29 2	1.41 (0.71, 2.11)	0.0002	–
Excellent vs GH29 3	1.06 (0.36, 1.76)	0.0035	–
Excellent vs GH29 4	1.15 (0.57, 1.73)	0.0002	–
Excellent vs GH29 5	0.95 (0.42, 1.48)	0.0007	–
EORTC GH30			0.939
Excellent vs very poor	−0.04 (−1.15, 1.06)	0.9378	–
Excellent vs GH30 2	1.48 (0.96, 2.00)	<0.0001	–
Excellent vs GH30 3	1.14 (0.62, 1.65)	<0.0001	–
Excellent vs GH30 4	1.45 (0.98, 1.92)	<0.0001	–
Excellent vs GH30 5	0.79 (0.37, 1.21)	0.0004	–
Excellent vs GH30 6	0.50 (0.09, 0.92)	0.0172	–

Reference groups for known validators were "never" for the PGIS severity group and "excellent" for the EORTC QLQ-C30 GHS/QoL items groups. EORTC GH29 and EORTC GH30 response options were rated using a 7-point scale ranging from very poor (1) to excellent (7).

- The PGIC anchor-based MWPC deterioration threshold and the average distribution-based clinical significance threshold triangulated well
  - The PGIC anchor-based threshold median estimate was 0.57 and the distribution-based clinical significance threshold mean estimate was 0.47, yielding an average of 0.52
  - Discrimination between "deterioration," "improvement," and "no change" based on the proposed deterioration threshold of 0.57 from the PGIC anchor was demonstrated by noticeable separation in eCDF and ePDF curves (**Figure 3A** and **B**)

Figure 3: eCDF (A) and ePDF (B) for the MySIm-Q symptom score



Total N is 36. Vertical reference lines are meaningful within-patient change; the solid line indicates anchor-based meaningful deterioration (0.57), while the dashed line is the average distribution-based threshold (0.47). 40.0% of the deterioration group met the threshold for meaningful deterioration, while 50.0% of the no change group did. 60.0% of the deterioration group met the distribution-based threshold, while 50.0% of the no change group did.

References

1. Jordan K, et al. *Support Care Cancer* 2014;22:417-26. 2. Ramsenthaler C, et al. *Eur J Haematol* 2016;97:416-29. 3. Mols F, et al. *Eur J Haematol* 2012;89:311-9. 4. Gries KS, et al. *Value Health* 2021;24:1807-19. 5. Hillengass J, et al. *Blood* 2023;142(suppl 1):1021. 6. Cohen AD, et al. *Blood* 2023;141:219-30. 7. Cronbach LJ. *Psychometrika* 1951;16:297-334. 8. Shrout PE, Fleiss JL. *Psychol Bull* 1979;86:420-8. 9. Campbell DT, Fiske DW. *Psychol Bull* 1959;56:81-105.

Multiple Myeloma

Presented by KS Gries at the International Society for Pharmacoeconomics and Outcomes Research Conference; May 5–8, 2024; Atlanta, GA, USA