#PRO142

Psychometric Performance of the Indolent Systemic Mastocytosis Symptom Assessment Form (ISM-SAF)

Alan L. Shields,[1] Fiona Taylor,[1] Roger E. Lamoureux,[1] Brad Padilla,[1] Kas Severson,[1] Tanya Green,[2] Brenton Mar[2]

[1]Adelphi Values, Boston, MA USA [2]Blueprint Medicines Corporation, Cambridge, MA

INTRODUCTION

- Systemic mastocytosis (SM) is a rare condition characterized by neoplastic cell growth in different organs, including bone marrow, skeletal system, lymph nodes, liver, spleen, and gastrointestinal tract.[1-3] It manifests as indolent SM (ISM), smoldering SM (SSM), and advanced SM (AdvSM).
- As a patient-reported outcome (PRO) questionnaire intended for use in regulated clinical trials, the Indolent Systemic Mastocytosis Symptom Assessment Form (ISM-SAF©, Blueprint Medicines Corporation) is an electronic daily diary to assess 12 signs and symptoms related to ISM and SSM.
- Though primarily developed for evaluating treatment efficacy hypotheses, the ISM-SAF can also be used to screen participants into (or out of) future clinical studies based on a minimum level of sign and symptom severity.

OBJECTIVE

- To psychometrically evaluate the ISM-SAF scores among patients with ISM and SSM and provide evidence that the ISM-SAF is "fit for purpose" for assessing treatment efficacy and establishing new product labeling claims.

METHODS

Study design

- Data were collected in February and March 2018 through a prospective, non-interventional study utilizing participants in the United States (US) diagnosed with ISM or SSM.
- Eligible participants were adults (i.e., at least 18 years of age, except in Alabama and Nebraska [≥19 years of age]) who self-reported a diagnosis with ISM or SSM.
  - Although not required, participants were asked to provide medical documentation confirming their diagnosis.
  - Patients who had any other hematologic malignancies/ blood cancers were not eligible for inclusion.
- Subjects completed assessments using a web-based, Health Insurance Portability and Accountability Act (HIPAA)-compliant platform (SurveyMonkey®) over the course of 15 days.
- A summary of the ISM-SAF, as well as other PRO assessments that were completed by participants, is presented in **Table 1**.
- As a daily diary, the ISM-SAF was completed on study days 1-15 and can be scored at an individual symptom or item level or to create a Total Symptom Score (TSS, Items 1-10 and 12), a Gastrointestinal Symptom Score (GSS, Items 2-3 and 12), and a Skin Symptom Score (SSS, Items 4-6).
- At a domain level, daily scores are created by summing the relevant items, weekly scores are created by averaging the daily scores over a 7-day period (minimum 4 days required), and bi-weekly scores are created by averaging scores over a 14-day period (minimum 7 days required).

Analysis populations

- Cross-sectional analysis population (CS-AP): All participants with ISM-SAF data at Day 1 (i.e., a daily TSS could be calculated) and sufficient Day 2-15 data to create a bi-weekly score (i.e., seven or more completed item scores from Day 2 to 15).
- Analyses were also conducted on a subsample of the CS-AP participants who had a confirmed ISM/SSM diagnosis (based on a review of the medical documents provided by the participant).
- Test-retest analysis population (TRT-AP): Participants who exhibited no change in PGIS from Day 1 to Day 15.

Psychometric evaluation

- **Internal consistency reliability**, which reflects the extent to which individual items from a multi-item scale are measuring the same general concept,[8] was investigated via Cronbach's alpha coefficient (α, range 0 to 1). α was calculated for the TSS, GSS, and SSS and again with each item removed to assess the impact that removal has on the overall α.
- **Test-retest reliability** assesses whether a measurement produces stable scores when administered under similar conditions at different timepoints during which no or minimal change in the patient's condition is expected. Using the TRT-AP, test-retest reliability for the TSS, GSS, and SSS was assessed using the intra-class correlation coefficient (ICC)[9] and its 95% confidence internal (CI). For item scores, test-retest reliability was examined using weighted kappa coefficients and the same weekly scores.[9]
- **Construct-related validity** is concluded upon evidence that scores produced by a target questionnaire relate to scores from other assessments in ways that are logical and according to *a priori* hypotheses.[10] The relationships between ISM-SAF scores and those generated by the supplementary assessments were examined via correlational analysis.
- **Known-groups analysis** was conducted to characterize the degree to which the ISM-SAF scores could distinguish among clinical groupings defined by PGIS responses, as well as SF-12v2® and MC-QoL tertiles.

RESULTS

Patient sample

- A total of 116 eligible patients were screened into the study and 103 were included in the CS-AP, including 58 with a clinically confirmed diagnosis of ISM (n=56, 96.6%) or SSM (n=2, 3.4%).
- In the CS-AP, mean age was 50.2 years (SD=12.6), 81.6% were female, and 98.1% were white.
- Demographic characteristics for the subsample with a confirmed diagnosis were similar.

ISM-SAF scores

- Participants reported symptom severity across the range of ISM-SAF response options (0–10), with responses tending to cluster near the lower end of the scale (i.e., less severe symptom experience).
- In the CS-AP, the highest mean biweekly scores were for items assessing fatigue (4.6), brain fog (3.2), and spots (3.1).
- Descriptive statistics for the subsample with a confirmed diagnosis were similar.

Table 1. Study assessments

Instrument	Administration schedule	Concepts assessed	Recall period	Response scale
ISM-SAF	Daily (Day 1-15)	Severity at worst: bone pain, abdominal pain, nausea, spots, itching, flushing, fatigue, headache, dizziness, brain fog, and diarrhea Frequency: diarrhea	Past 24 hours	11-point NRS (higher scores = worse)
Patient Global Impression of Severity (PGIS)	Days 1 and 15	Overall severity of ISM or SSM symptoms at a given timepoint	Global assessment (present)[a]	Five-point verbal rating scale (VRS)
12-Item Short Form Survey, Version 2 (SF 12v2®)[4,5]	Days 1 and 15	Physical and emotional health and related functional limitations	Past week[b]	Five-point and three-point VRS
Mastocytosis Quality of Life Questionnaire (MC-QoL)[6]	Days 1 and 15	Health-related quality of life impairment in patients with cutaneous mastocytosis and ISM (symptoms, emotions, social life/functioning, and skin)	Past two weeks[c]	Five-point VRS
Functional Assessment of Cancer Therapy – Cognitive Function (FACT-Cog)[7]	Day 1	Cognitive impairment and impact on one's daily life	Past seven days	Five-point scale (scored 0–4)

[a]ISM-SAF daily scores used for psychometric analyses at Day 15 to match PGIS recall period
[b]ISM-SAF weekly scores (Days 9-15) used for psychometric analyses at Day 15 to match SF 12v2® recall period
[c]ISM-SAF biweekly mean scores (Days 2-15) used for psychometric analyses at Day 15 to match MC-QoL recall period

RESULTS (continued)

Internal consistency reliability (Table 2)

- For the CS-AP TSS Day 15 (biweekly score), α was 0.884; for the confirmed diagnosis subsample TSS at Day 15 (biweekly score), α was 0.876. All alphas were greater than or equal to 0.67 for both analysis populations, indicating sufficient internal consistency among items.
- Removal of items typically reduced overall alpha coefficients; any instances in which alpha increased (e.g., Item 4, spots) were only marginal.

Test-retest reliability (Table 3)

- Test-retest reliability estimates comparing week 1 (an average of scores generated on Day 2 to 8) and week 2 (an average of scores generated on Day 9 to 15) were all acceptable (≥0.87) at both the domain (ICC) and item levels (weighted kappa).

Construct-related validity (Table 4)

- The relationships between the TSS and other variables were strong and in the expected direction:
  - TSS scores were more strongly correlated with variables assessing symptoms and physical function (such as the role physical and bodily pain domains of the SF 12v2® and the symptoms domain of the MC-QoL) and less strongly correlated with variables associated with more distal disease impacts (such as the mental component score or the role emotional domain of the SF 12v2®).
  - Participants reporting increased symptom involvement on the ISM-SAF also rate themselves as more severely afflicted on the PGIS.
- Correlations with other measures were generally greater for the TSS than for the GSS and SSS, except for the MC-QL Skin domain, which correlated most strongly with the SSS as expected.

Known-groups analysis (Table 5)

- Based on results from the CS-AP and the confirmed diagnosis subsample, and in each of the three clinical groupings, TSS, GSS, and SSS scores were clearly distinct, in the hypothesized direction (i.e., participants with greater symptoms and impacts, as assessed by the PGIS, MC-QoL, and SF 12v2®, also scored higher on the ISM-SAF), and those differences were statistically significant ($p<0.05$).

Table 2. Internal consistency reliability (α) on the biweekly ISM-SAF total symptom scale and domain scores

Domain	CS-AP (n=103)	Confirmed diagnosis (n=58)
TSS	**0.884**	**0.876**
α if item deleted	--	--
Item 1: Bone pain	0.870	0.862
Item 2: Abdominal pain	0.866	0.859
Item 3: Nausea	0.870	0.861
Item 4: Spots	0.896	0.881
Item 5: Itching	0.875	0.866
Item 6: Flushing	0.870	0.859
Item 7: Fatigue	0.861	0.849
Item 8: Dizziness	0.868	0.859
Item 9: Brain Fog	0.876	0.867
Item 10: Headache	0.871	0.861
Item 12: Diarrhea severity	0.883	0.887
GSS	**0.777**	**0.685**
SSS	**0.667**	**0.700**

Table 3. Test-retest reliability between Weeks 1 and 2 on Patient Global Impression of Severity stable participants (TRT-AP; n=61)

	ICC or K(w) (95%CI)
TSS	**0.962 (0.936-0.977)**
GSS	**0.936 (0.894-0.962)**
SSS	**0.962 (0.937-0.977)**
Item 1: Bone pain	0.943 (0.905-0.966)
Item 2: Abdominal pain	0.922 (0.870-0.953)
Item 3: Nausea	0.937 (0.895-0.962)
Item 4: Spots	0.974 (0.957-0.985)
Item 5: Itching	0.902 (0.837-0.941)
Item 6: Flushing	0.971 (0.952-0.983)
Item 7: Fatigue	0.951 (0.918-0.971)
Item 8: Dizziness	0.929 (0.881-0.957)
Item 9: Brain fog	0.956 (0.926-0.973)
Item 10: Headache	0.905 (0.841-0.943)
Item 11: Diarrhea (frequency)	0.885 (0.809-0.931)
Item 12: Diarrhea	0.869 (0.781-0.921)

Table 4. Spearman correlations of ISM-SAF total and domain scores with other measures administered at Day 15 ≥0.6=green <0.3=red

Concurrent measure	Cross-sectional analysis population (N=103)			Confirmed diagnosis subsample (n=58)		
	TSS	GSS	SSS	TSS	GSS	SSS
SF-12: Physical Functioning	-0.585	-0.480	-0.265	-0.685	-0.530	-0.484
SF-12: Role Physical	-0.741	-0.608	-0.390	-0.729	-0.547	-0.528
SF-12: Bodily Pain	-0.722	-0.557	-0.418	-0.760	-0.514	-0.585
SF-12: General Health	-0.560	-0.417	-0.329	-0.667	-0.432	-0.511
SF-12: Vitality	-0.504	-0.441	-0.212	-0.453	-0.305	-0.222
SF-12: Social Functioning	-0.584	-0.568	-0.317	-0.577	-0.505	-0.408
SF-12: Role Emotional	-0.502	-0.435	-0.307	-0.459	-0.377	-0.316
SF-12: Mental Health	-0.611	-0.553	-0.457	-0.583	-0.450	-0.499
SF-12: Physical Component Score	-0.631	-0.493	-0.308	-0.725	-0.511	-0.526
SF-12: Mental Component Score	-0.483	-0.465	-0.346	-0.425	-0.356	-0.315
MC-QoL: Symptoms	0.832	0.676	0.486	0.833	0.620	0.601
MC-QoL: Social Life/Functioning	0.773	0.625	0.506	0.768	0.547	0.604
MC-QoL: Emotions	0.712	0.580	0.512	0.710	0.493	0.727
MC-QoL: Skin	0.635	0.459	0.779	0.661	0.397	0.795
MC-QoL: Total Score	0.849	0.679	0.587	0.853	0.602	0.730
PGIS	0.618	0.454	0.446	0.610	0.373	0.543

Table 5. Known groups analysis of the ISM-SAF total and domain scores based on PGIS, MC-QoL, and SF-12v2® assessments administered at Day 15

PRO	Group	Cross-sectional analysis population (N=103)				Confirmed diagnosis subsample (n=58)			
		n	TSS M (SD)	GSS M (SD)	SSS M (SD)	n	TSS M (SD)	GSS M (SD)	SSS M (SD)
PGIS	Absent/Minimal	41	16.5 (14.8)	3.0 (4.8)	5.3 (4.2)	26	18.5 (14.1)	3.9 (5.1)	4.7 (3.9)
	Moderate	43	29.3 (12.5)	5.6 (3.9)	9.2 (5.4)	22	32.4 (13.0)	5.8 (3.7)	10.3 (5.3)
	Severe/Very Severe	18	48.3 (19.6)	9.6 (7.4)	12.2 (7.0)	9	50.4 (20.7)	9.3 (7.7)	11.8 (5.7)
MC-QoL	Mild	37	13.4 (8.2)	2.3 (2.4)	5.2 (4.3)	23	16.6 (9.9)	3.4 (3.0)	5.4 (4.7)
	Moderate	32	27.9 (8.7)	5.0 (3.1)	9.1 (5.0)	15	29.5 (9.3)	5.1 (3.3)	9.5 (5.4)
	Severe	33	42.2 (13.9)	9.2 (4.6)	10.9 (4.7)	19	42.0 (12.0)	8.3 (3.6)	11.0 (3.3)
SF-12v2®	Mild	34	17.8 (15.6)	3.1 (3.5)	7.0 (6.2)	19	17.2 (11.5)	3.1 (2.6)	5.6 (4.4)
	Moderate	33	23.9 (10.2)	4.7 (3.3)	7.2 (4.3)	18	24.8 (8.7)	4.7 (3.5)	8.1 (5.3)
	Severe	34	40.2 (15.0)	8.3 (5.5)	10.8 (4.7)	19	43.9 (13.4)	8.8 (4.2)	11.7 (3.9)

CONCLUSIONS

- These psychometric results support the conclusion that the ISM-SAF can produce trustworthy scores when administered to patients in the target population in that the domain and item scores are reliable, construct-valid, and able to distinguish among clinically unique groups.
- Further evaluation of the psychometric properties of the ISM-SAF is planned using clinical trial data to provide additional evidence that the ISM-SAF is "fit for purpose" for assessing efficacy and establishing labeling claims.

FUNDING / DISCLOSURES

- This study was funded by Blueprint Medicines Corporation. Blueprint Medicines Corporation participated in the interpretation of data, review, and approval of the publication.
- BM and TG are employees of Blueprint Medicines Corporation and ALS, FT, REL, BP, and KS are employed by Adelphi Values, which received payment from Blueprint Medicines Corporation for participation in this research.

REFERENCES

1. Kluin-Nelemans HC et al. Cladribine therapy for systemic mastocytosis. Blood. 2003;102(13):4270-4276.
2. Valent P, et al. Aggressive systemic mastocytosis and related mast cell disorders: current treatment options and proposed response criteria. LeukRes. 2003;27(7):635-641.
3. Zanotti R et al. Isolated bone marrow mastocytosis: an underestimated subvariant of indolent systemic mastocytosis. Haematologica. 2011;96(3):482-484.
4. Maruish ME ed User's Manual for the SF-36v2®. 3rd ed. Lincoln, RI: QualityMetric Inc; 2011.
5. Ware JE, Jr. et al. 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220-233.
6. Siebenhaar F et al. Development and validation of the mastocytosis quality of life questionnaire: MC-QoL. Allergy. 2016;71(6):869-877.
7. Wagner LI et al. Measuring patient self-reported cognitive function: development of the functional assessment of cancer therapy-cognitive function instrument. J Support Oncol. 2009;7(6):W32-W39.
8. Thompson B. Understanding reliability and coefficient alpha, really. In: Thompson B, ed. Score reliability: contemporary thinking on reliability issues. Thousand Oaks, CA: Sage Publications, Inc.; 2003:3-21.
9. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. PsycholBull. 1979;86(2):420-428.
10. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009..

ISPOR Europe 2019, Copenhagen, Denmark, 2-6 Nov 2019