Psychometric Evaluation of the Mastocytosis Symptom Severity Daily Diary (MS2D2) Instrument in Patients with Nonadvanced Systemic Mastocytosis (NonAdvSM)

James C. Marcus, PhD[1] , Cem Akin, MD[2] , Jenna Zhang, PhD[3] , Rachael Easton, MD, PhD[3] , Michelle Lim-Watson, PhD, MPH, MBA[4] , Frank Siebenhaar, MD[5], Ralph Turner, PhD[1]

[1]IQVIA, New York, NY, USA, [2]University of Michigan, Ann Arbor, MI, USA, [3]Cogent Biosciences, Waltham, MA, USA, [4]MCPHS University, Boston, MA, USA, [5]Insititute of Allergology, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin and Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Berlin, Germany

Summit NONADVANCED SM

Poster #PCR173

INTRODUCTION

- Systemic mastocytosis (SM) is a heterogeneous disease characterized by the accumulation of neoplastic mast cells in the bone marrow and other extracutaneous tissues. It includes subtypes such as advanced SM and NonAdvSM (Indolent, smoldering, and bone marrow), each presenting a range of symptoms from the release of mast cell mediators.[1-2]
- The MS2D2 is a de novo 17-item disease-specific daily diary presented electronically that evaluates symptom severity in patients with various subtypes of NonAdvSM.
- The objective of this study was to assess the psychometric properties of the MS2D2 in Part 1 (n=54) of the Summit study (NCT05186753): A multi-part, randomized, double-blind, placebo-controlled, study to evaluate bezuclastinib in patients with moderate to severe NonAdvSM.

METHODS

- **MS2D2 Item Response Scales:** Severity items (15 of the 17) used a 0-10 numerical rating scale (NRS); diarrhea frequency used a 'spinner'; and mast cell reaction was binary (Yes/No).
- **MS2D2 Total Symptom Score (TSS) and Domains:** Using 11 items, MS2D2 yields an item-weighted TSS and four domains (# of items): Dermatological (4), Other [GI/Pain] (4), Neurocognitive (2), and Fatigue (1). Daily scores are averages of items (0-10 range; TSS presented as a 0-110 sum); 14-day scores (primary unit of analysis) average daily scores over 14 days.
- Part 1 Summit study data investigated:
  - **Structural validity** (via daily score at baseline):
    - **Item response distributions**
    - **Inter-item correlations** (Pearson)
    - **Confirmatory factor analysis (CFA)** to test and refine hypothesized domain and TSS structures
  - **Reliability** (14-day scores)
    - **Internal consistency** via Cronbach's α and overlap corrected item-total correlations
    - **Test-retest** between screening and baseline on stable patients using agreement-based Intraclass Correlation Coefficient (ICC)
  - **Construct validity** (14-day scores)
    - **Convergent** via Pearson correlation
    - **Known groups** via ANOVA (Analysis of Variance) and Helmert contrast
  - **Sensitivity to change** via polyserial correlations
  - **Clinically meaningful change thresholds** via anchor-based approach (Patient Global Impression of Severity (PGIS) change and Patient Global Impression of Change (PGIC))

RESULTS

Sample

- Analysis was conducted on the 54 intent-to-treat (ITT) patients of Part 1 of the Summit trial. (only 51 with MS2D2 data).

Table 1 | Patient Demographics and Clinical Characteristics at Eligibility (N = 54)

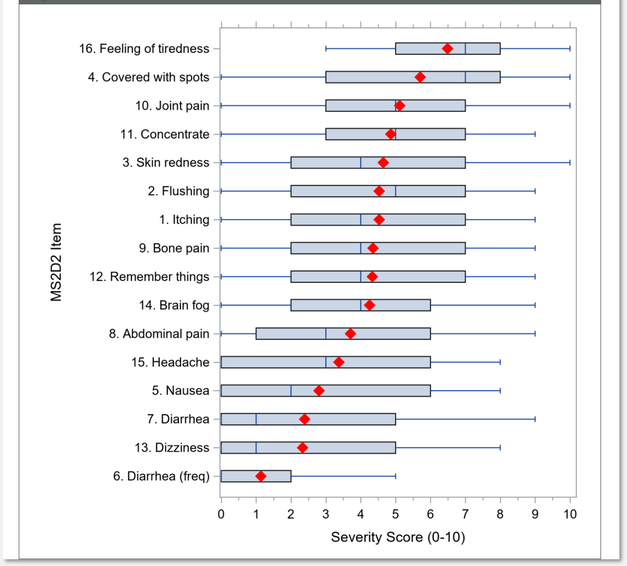
Characteristic	Statistic
Female, n (%)	36 (66.7)
Age: Median (Interquartile Range [IQR]; Range)	51 (41-65; 27-76)
NonAdvSM, n (%)	
Indolent (including BMM)	51 (94.4)
Smoldering	3 (5.6)
Mastocytosis Activity Score (MAS) Total Score[3]: Median (IQR; Range)	43.4 (36.1-50.0; 26.3-71.6)
Mast Cell Burden: Median (IQR; Range)	
Serum Tryptase (ng/mL)	44.3 (23.2-78.8; 9.2-592.0)
Bone marrow mast cell (%)	15 (5-25; 1-80)
KITD816V mutation allele burden (%)	0.14 (0.03-2.26; 0-32.48)

RESULTS, continued

Item Response Distributions

- Item distributions at Baseline (Figure 1) indicate patients used the entire 0-10 response scale, with most symptoms averaging severity scores of 4 to 5.
  - Tiredness was the most severe symptom (mean 6.5, median 7.0, showing low variability) followed by 'covered with spots' (mean 5.7, median 7.0, with higher variability).
  - Diarrhea and dizziness have lower severity scores, with means of ~2 and medians of 1.

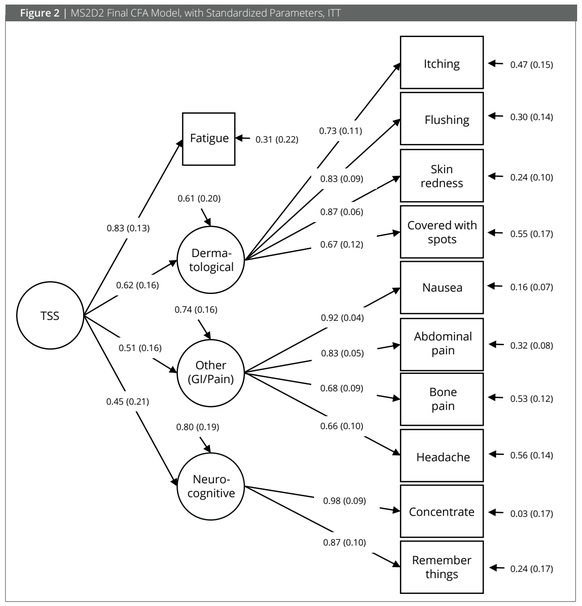
Figure 1 | Box plot of MS2D2 Item Distributions (Continuous) at Baseline (D-1), Ordered by mean severity, ITT



Confirmatory Factor Analysis

- Structural validity analyses (2nd order CFA) supported an 11-item MS2D2 with four domains, Dermatological, Neurocognitive, Other (GI/Pain), and Fatigue, all loading on the TSS.

Figure 2 | MS2D2 Final CFA Model, with Standardized Parameters, ITT



MS2D2 Inter-item Correlations

Table 2 | Inter-item Pearson Correlations (r) for MS2D2 Items at Baseline Day (D-1), ITT, N = 51

Intended Domain	#	Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Dermatological	1	Itching	-																
	2	Flushing	**.58**	-															
	3	Skin redness	**.61**	**.74**	-														
	4	Covered with spots	**.65**	**.53**	**.59**	-													
Gastrointestinal	5	Nausea	**.14**	.40	.20	.07	-												
	6	Diarrhea (freq)	.33	.30	.16	.25	.12	-											
	7	Diarrhea worst	**.50**	**.45**	.30	.43	.35	**.74**	-										
Pain	8	Abdominal pain	.15	.29	.03	.10	**.76**	.07	.33	-									
	9	Bone pain	.23	.27	.25	.13	**.61**	.11	.29	**.57**	-								
	10	Joint pain	.24	.39	.38	.15	**.57**	.19	.37	.33	**.69**	-							
Neurocognitive	11	Concentrate	.11	.22	.18	.12	.35	.08	.16	.22	.22	.36	-						
	12	Remember things	.09	.23	.19	.04	.37	.09	.07	.22	.28	.38	**.86**	-					
	13	Dizziness	.32	.39	.29	.21	**.45**	.08	.24	.37	**.46**	**.51**	**.48**	**.52**	-				
	14	Brain fog	.10	.19	.17	-.03	.24	.03	.02	.14	.16	.33	**.78**	**.78**	**.52**	-			
	15	Headache	.12	.27	.19	.08	**.61**	.00	.18	**.55**	**.45**	.27	.15	.13	.41	.19	-		
	16	Feeling of tiredness	.38	**.49**	**.50**	.17	.38	.15	.27	.30	.41	**.48**	.36	.30	.20	.23	.19	-	
	17	Mast cell reaction	.23	.09	.23	.27	.06	.11	.14	.21	.12	-.04	-.06	-.13	.09	.04	.24	.03	-

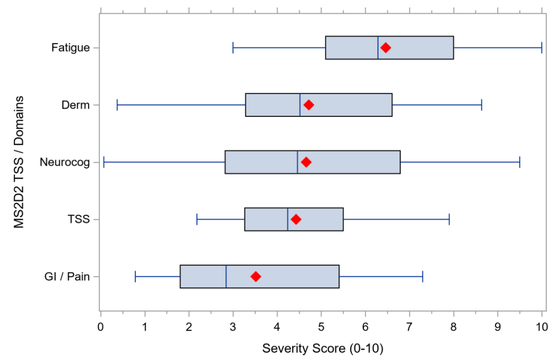
Note(s): Baseline day denotes the last diary entry prior to start of treatment
bold denotes *r* between .45 and .75; **red** denotes r>.75

- **Items had inter-item correlations at Baseline (Table 2) mostly in line with expected domains.**
  - Dermatological items (1-4) most clearly suggested a distinct domain.
  - GI and Pain scales combined due to high *r* between nausea and pain items (and low *r* with diarrhea [7]).
  - A subset of intended Neurocognitive items (11,12,14) had high Pearson's correlation coefficient (*r* >.75), suggesting grouping.
    - Headache assigned to Pain given higher *r* with those items (and nausea).
    - Tiredness treated as a separate single item Fatigue domain due to low r but high severity.
    - Dizziness excluded due to high cross-loadings with Pain (and low severity).

Scale Level Descriptive Statistics

- Baseline descriptive statistics revealed that TSS scores (0-10) fell near center of the scale (mean 4.4; median 4.2), with narrow IQR (3.3 to 5.5) reflecting consistent experiences across patients.
- Fatigue was the most severe domain and Other (GI/Pain) the least, with domains having ~4-pt IQR spread and 6-pt 95-percentile spread.

Figure 3 | Box plot of MS2D2 TSS and Domain Response Distributions (0-10 scale) at Baseline, ITT, N = 51



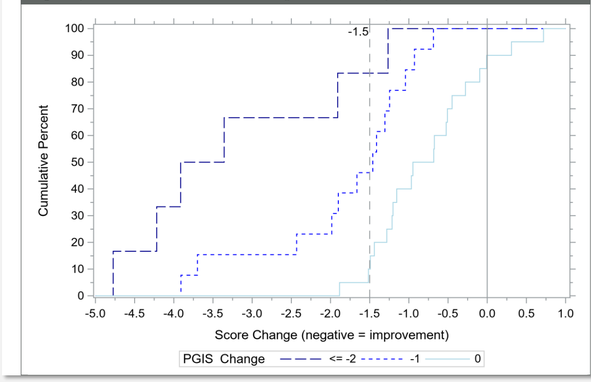
Scale-level Measurement Properties

- **Internal Consistency:** MS2D2 scales had high Cronbach's alpha values (range 0.84 - 0.92) at baseline, supporting internal consistency.
- **Test-retest Reliability**: MS2D2 Scales demonstrated excellent test-retest reliability (ICC range: 0.86 - 0.97) from screening to baseline.
- **Convergent Validity:** MS2D2 TSS showed strong correlations with other disease-specific measures, e.g., MAS Total score (r=0.90) and MC-QoL Total score (r=0.57)
- **Known-Groups Validity**: MS2D2 scales effectively distinguished between varying symptom severities, with significant differences across PGIS levels ($p < 0.05$), e.g., TSS (0-10 scale) means increased from 2.78 (Mild) to 5.65 (Very Severe).
- **Sensitivity to Change**: MS2D2 TSS demonstrated moderate to strong correlations between change over time in patient impressions symptom severity (PGIS, r=.71) and impression of change (PGIC, r=-0.55).

Within-patient Clinically Meaningful improvement Thresholds:

- MS2D2 TSS and domains support a ~ 1.5 – 2.0 raw change band (0-10 scale), which corresponds to a 34-45% change from baseline.
- Distribution-based (MDC) analyses suggested any change over ~1 would be distinguishable from measurement error.
- eCDFs at Cycle 4 supported MMRM anchor-based analyses
  - E.g., For TSS, PGIS change (Figure 4) suggested meaningful within-patient change (MWPC) thresholds of 1.5, with ~45% of the '1-pt improvement ' group exceeding the threshold vs. only about 5% of those in the 'No change' (0) group.

Figure 4 | CDF for MS2D2 (0-10 scale) with PGIS Change as Anchor – Total Symptom Score



CONCLUSIONS

Part 1 (N = 51) Summit study data support the MS2D2 as a reliable and valid instrument for capturing the NonAdvSM patient experience.

MS2D2 demonstrated:

- Structural validity of MS2D2 domains and TSS
- High internal consistency and test-retest reliability.
- Strong construct validity, with high correlations with MAS and MC-QoL and scores distinguishable by different PGIS levels.
- Strong ability to detect change, with MWPC thresholds in a ~1.5 to 2.0 band on a 0-10 scale.

Results provide evidence for the use of MS2D2 in clinical trials.

International Society for Pharmacoeconomics and Outcomes Research (ISPOR) annual meeting | **May 13-16 in Montreal, QC, Canada**

References: 1. Pardanani, A. (2019). Systemic mastocytosis in adults: 2019 update on diagnosis, risk stratification and management. American Journal of Hematology, 94(3), 363–377 2. Ustun, C., Arock, M., Kluin-Nelemans, et al., Advanced systemic mastocytosis: From molecular and genetic progress to clinical practice. Haematologica, 101(10), 1133–1143. 3. Siebenhaar et al., Development and validation of the mastocytosis activity score. Allergy 2018, 73 (7), 1489–1496.

Acknowledgments/Disclosures This study was funded by Cogent Biosciences. The authors would like to acknowledge the patients who participated in the trial and those who support them, including family, caregivers, and patient advocates.