Validation of the labeled magnitude scale for the assessment of itch intensity in patients with chronic pruritus

Claudia Zeidler, MD, Marvin Röchter, MD, Sonja Ständer, MD, Manuel P. Pereira, MD, PhD*

Introduction: The visual analogue scale (VAS), the numerical rating scale (NRS), and the verbal rating scale (VRS) are routinely used to assess pruritus intensity. However, these scales have limitations, especially the ceiling effect of linear scales (VAS, NRS) and the reduced sensitivity to change of the VRS. In the labeled magnitude scale (LMS) consisting of a vertical line with verbal anchors distributed quasi-logarithmically, the ceiling effect is reduced, allowing a higher discriminative assessment at the higher end of the scale.

Methods: A total of 172 patients completed the LMS twice within 30–60 minutes to assess the reliability of the scale, as well as the NRS and VAS for analyses of convergent validity. Sensitivity to change of the LMS was investigated by analyzing changes in scores after a twice-daily application of an emollient containing menthoxypropanediol for 14 days.

Results: Test-retest reliability was excellent for the LMS assessing average (Cronbach α: 0.955) and worst (Cronbach α: 0.945) pruritus intensity in the last 24 hours, while strong to very strong correlations were observed between the LMS and NRS assessing the worst (r = 0.783) and average (r = 0.808) pruritus intensity, respectively. Treatment with an emollient lead to a significant decrease in LMS scores (P < 0.01). In patients with severe pruritus at baseline (VRS ≥ 3), we recorded a 30.2% improvement of average pruritus intensity using the LMS compared with a 25.0% improvement using the NRS and a 27.8% improvement of worst pruritus intensity using the LMS compared with an 11.1% improvement with the NRS. Most patients considered the LMS an appropriate instrument to assess pruritus intensity (89.6%) and would use it again (91.9%).

Discussion: The LMS is an appropriate well-accepted instrument to assess CP. It is especially useful to detect variations in pruritus intensity in patients with severe CP.

Keywords: Labeled magnitude scale, Numerical rating scale, Visual analogue scale, Verbal rating scale, Chronic pruritus, Itch intensity

Chronic pruritus (CP), defined as pruritus lasting for 6 weeks or longer, is a subjective multidimensional symptom, which is complex to measure[11]. In routine dermatological care and clinical trials the use of tools measuring the intensity of pruritus has been prioritized[2-3]. Linear instruments, such as the numerical rating scale (NRS; range: 0–10) and the visual analogue scale (VAS; range: 0–10), and the ordinal verbal rating scale (VRS; range: 0–4) have been validated to assess CP and are widely used[4-5]. However, these scales show important shortcomings. One limitation of the linear scales (NRS, VAS) is the ceiling effect, in which distinct intensities at the higher end of the scale cannot be adequately discriminated[6], resulting in an inaccurate assessment of severe and very severe pruritus. As for the VRS, the small number of categories of this scale limits its sensitivity to change[7], especially when comparing to the 11-point NRS[8].

To address these issues, we adapted the labeled magnitude scale (LMS) for the assessment of CP (Fig. 1). LMS is a tool developed for assessing sensory symptoms consisting of a 10 cm long, vertical line with verbal anchors distributed in a quasi-logarithmic manner[9]. It contains thus elements of the VRS (the presence of verbal anchors) while allowing patients to mark the perceived pruritus intensity on a continuous line, as in the VAS. Importantly, due to its architecture, the ceiling effect is reduced in the LMS compared with the NRS or VAS. The LMS has been validated for the assessment of acute and chronic pain[10]. In addition, it has been used in acute pruritus[11]; however, the validation for the measurement of CP is still lacking.

Aim of this study was to validate the LMS in German for the assessment of CP in regard to its test-retest reliability and convergent validity, as well as assess the sensitivity to change of this novel instrument in CP.

Methods

Patients

Adult patients with CP presenting at the Center for Chronic Pruritus, Department of Dermatology of the University Hospital Münster, Germany.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

*Corresponding author. Address: Department of Dermatology and Center for Chronic Pruritus, University Hospital Münster, Von-Esmarch-Str. 58, Münster 48149, Germany. Tel: +49 251 83 57470; fax: +49 251 83 52559. E-mail address: manuelepereira@ukmuenster.de (M.P. Pereira).

Copyright © 2021 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The International Forum for the Study of Itch. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Itch (2021) 6:e52

Received 29 December 2020; Accepted 25 April 2021
Published online 13 September 2021
http://dx.doi.org/10.1097/ITX.000000000000052
Münster (Germany) were consecutively enrolled. Only patients showing dry skin and reporting in the previous 24 hours a maximal itch intensity of 3 or more on the NRS were considered. Exclusion criteria included acute exacerbation of a skin condition, cutaneous infection, pregnancy or breastfeeding, and drug abuse. All study procedures were performed respecting the Helsinki declaration and later revisions. The study was approved by the local Ethics Committee (Medical Faculty of the University of Münster, 2017-415-f-S) and was registered at the German Registry of Clinical Trials (DRKSDRKS00013074).

**Study design**

This study consisted of 2 parts. In the first part, the feasibility of the use of the LMS in the clinical routine was assessed. Patients with CP (n = 120) filled out the scale before a routine appointment. Generally, the scale was understood and well accepted by the patients (data not shown), so that validation of the scale was sought in the subsequent phase of this study. The second part of the study consisted of 2 visits. In the first visit, CP patients were asked to complete the LMS and the NRS twice in an interval of 30–60 minutes, as well as the other pruritus intensity scales (VAS, VRS) once. Afterward, patients were instructed to use an emollient containing menthoxypipanediol twice daily until the next study visit 2 weeks later. In the second visit, patients completed once more the LMS, NRS, VAS and VRS, and a feasibility questionnaire. Both the average and the worst pruritus intensity of the previous 24 hours were assessed with the LMS. Table 1 summarizes the instruments used in this study for the assessment of pruritus intensity, including recall periods.

**Reliability and validation**

To assess the test-retest reliability of the LMS, patients were asked to complete this scale twice within an interval of 30–60 minutes (study visit 1). Convergent validity was evaluated by correlating LMS scores to pruritus intensity scores provided by external instruments, namely the NRS and VAS.

**Sensitivity to change**

In order to assess the sensitivity to change, LMS scores before and after treatment with an emollient containing menthoxypipanediol were compared in patients reporting an improvement of the pruritus. Change in LMS scores was compared with change in the NRS. To check whether using the LMS can diminish the ceiling effect, we analyze the course of the scores of LMS and NRS of those patients who indicated their pruritus intensity as severe or very severe on the VRS before treatment begins.

**Feasibility questionnaire**

After completing the LMS, patients were asked to fill out a feasibility questionnaire. Here it was questioned whether patients considered the LMS to be clearly formulated, if it was difficult to decide on a score, whether the verbal anchors were adequately distributed and whether the scale was too short. Finally, patient perspective on the suitability of the LMS for the assessment of CP and whether they would use this instrument again was assessed.

**Statistics**

We used SPSS v. 26.0 for Windows (IBM Corporation, Armonk, NY) for statistical analyses. Descriptive data are shown as number of patients (percentage), median [interquartile range (IQR)] and range. To assess test-retest reliability, intraclass correlation coefficients (ICC) and Cronbach α were used. Correlation coefficients between LMS and the external scales to assess convergent validity were calculated with the Spearman rank test. We performed the Wilcoxon test for analyses of sensitivity to change. The level of significance was set at $P \leq 0.05$.

**Results**

**Patients**

A total of 172 patients with CP completed the validation study. Demographic data, the clinical classification of patients according to their skin status and the etiology of the pruritus are shown in Table 2.

---

**Table 1**

| Scale | Assessed parameter | Recall period |
|-------|--------------------|---------------|
| LMS   | Average pruritus intensity | 24 h |
|       | Worst pruritus intensity | 24 h |
| NRS   | Average pruritus intensity | 24 h |
|       | Worst pruritus intensity | 24 h |
| VAS   | Average pruritus intensity | 4 wk |
|       | Worst pruritus intensity | 4 wk |
| VRS   | Average pruritus intensity | 24 h |

LMS indicates labeled magnitude scale; NRS, numerical rating scale; VAS, visual analogue scale; VRS, verbal rating scale.
**Table 2**

| Study population. | All (n = 172, 100%) | Female (n = 89, 51.7%) | Male (n = 83, 48.3%) |
|-------------------|---------------------|------------------------|----------------------|
| Age (y)           | Median [IQR]        | Range                  | Clinical classification, n (%) |
|                   | 62 [52–74]         | 18–91                  | I: lesion (inflamed skin) 53 (31.0) |
|                   |                    |                        | II: nonlesional skin 66 (38.6) |
|                   |                    |                        | III: chronic scratch lesions 52 (30.4) |
|                   |                    |                        | Etiological classification, n (%) |
|                   |                    |                        | I: dermatological 72 (41.9) |
|                   |                    |                        | II: systemic 14 (8.1) |
|                   |                    |                        | III: neurological 63 (36.6) |
|                   |                    |                        | IV: psychiatric 1 (0.6) |
|                   |                    |                        | V: multifactorial 17 (9.9) |
|                   |                    |                        | VI: unknown origin 5 (2.9) |

Sex and age distribution, as well as the clinical and etiological classification of chronic pruritus is shown for the study population. Data are shown as median [IQR], range and number of patients (percentage), as appropriate.

**IQR** indicates interquartile range.

---

**Test-retest reliability**

Test-retest reliability was excellent for the LMS measuring both worst and average pruritus intensity according to the Cronbach α, while ICCs revealed an almost perfect agreement for both scales. Similar reliability scores were observed for the NRS (Table 3).

**Convergent validity**

Strong to very strong correlations were observed between the LMS and NRS assessing the average pruritus intensity in the previous 24 hours \( (r = 0.808, P = 0 < 0.001, n = 169) \) and between LMS and NRS assessing the worst pruritus intensity in the previous 24 hours \( (r = 0.783, P = 0 < 0.001, n = 169) \). Moderate correlations were recorded for the correlations between the LMS and the VAS (Table 4).

**Sensitivity to change**

Comparing baseline and follow-up scores of the LMS and NRS a significant reduction of the intensity in both measuring instruments were found (Table 5). However, in patients with severe pruritus at baseline the LMS scores had a stronger reduction than the NRS scores (Table 6).

---

**Feasibility**

Most patients considered the LMS an appropriate instrument to assess pruritus intensity \( (n = 95/106, 89.6\%) \) and would use it again \( (n = 102/111, 91.9\%) \). For the majority of patients the LMS was clearly formulated \( (n = 105/114, 92.1\%) \) and the distribution of the verbal anchors was adequate \( (n = 107/114, n = 93.9\%) \). Only few patients found the LMS too short \( (n = 5/115, 2.9\%) \) or had difficulty in deciding on a score \( (n = 18/113, 15.9\%) \).

**Discussion**

We validated for the first time the LMS for the use in CP. In previous studies, the LMS was used in acute experimentally induced pruritus and proved to be more reliable than the VAS\(^{[11]}\). In our investigations, the LMS assessing the worst and average chronic pruritus intensity of the previous 24 hours showed a very good convergent validity with strong to very strong correlations with the NRS, while test-retest reliability was excellent for the LMS assessing both average and worst pruritus intensity in the last 24 hours. Moreover, this scale showed sensitivity to change, as improvement of itch after treatment with an emollient containing mentholpropanediol resulted in significant lower LMS.

---

**Table 3**

| Scale                  | Score—test (median [IQR]) | Score—retest (median [IQR]) | Cronbach α* | ICC\(^d\) (95% CI) |
|------------------------|---------------------------|-----------------------------|-------------|-------------------|
| Labeled magnitude scale (average pruritus, 24 h) | 3.3 [2.0–4.5], n = 170 | 3.2 [1.8–4.4], n = 81 | 0.955 | 0.912 (0.866–0.942) |
| Labeled magnitude scale (worst pruritus, 24 h) | 3.8 [2.6–5.3], n = 171 | 3.7 [2.6–5.3], n = 77 | 0.945 | 0.898 (0.843–0.934) |
| Numerical rating scale (average pruritus, 24 h) | 5.0 [4.0–7.0], n = 171 | 6.0 [4.0–7.0], n = 68 | 0.980 | 0.961 (0.938–0.976) |
| Numerical rating scale (worst pruritus, 24 h) | 7.0 [5.0–8.0], n = 170 | 7.0 [5.3–8.0], n = 68 | 0.981 | 0.963 (0.941–0.977) |

Labeled magnitude scale and numerical rating scale were completed twice by patients with chronic pruritus within 30–60 minutes. Cronbach α and ICC are shown for each instrument.

\*Cronbach α < 0.5: unacceptable, 0.5 ≤ α < 0.6: poor, 0.6 ≤ α < 0.7: questionable, 0.7 ≤ α < 0.8: acceptable, 0.8 ≤ α < 0.9: good, α ≥ 0.9: excellent.

\( ^d \)ICC < 0: less than chance agreement, 0.01–0.20: slight agreement, 0.21–0.40: fair agreement, 0.41–0.60: moderate agreement, 0.61–0.80: substantial agreement, 0.81–0.99: almost perfect agreement.

\( ^c \)ICc indicates confidence interval; ICC, intraclass correlation coefficients; IQR, interquartile range.
scores. Thus, the LMS is a suitable instrument to assess pruritus intensity in CP conditions, augmenting the spectrum of indications of the LMS.

The innovative aspect of the LMS is its ability to reduce the ceiling effect observed in linear scales (NRS, VAS), due to its quasi-logarithmic distribution of the verbal anchors. Indeed, in patients with severe or very severe pruritus, as it can detect variations of pruritus intensity at the higher end of the scale, which are not time in patients with severe pruritus, as it can detect variations of observed for the whole study population who suffered from mild of 11.1% using the NRS. However, this difference was not pruritus intensity using the LMS compared with an improvement of 25.0% using the NRS, and a 27.8% improvement of worst pruritus intensity using the LMS compared with an improvement of 11.1% using the NRS. However, this difference was not observed for the whole study population who suffered from mild to very severe pruritus intensity according to the VRS. As such, the LMS is particularly helpful to assess pruritus intensity over time in patients with severe pruritus, as it can detect variations of pruritus intensity at the higher end of the scale, which are not depicted by the NRS or VAS. These findings are in agreement with previous studies showing higher sensitivity of the LMS for the assessment of high intensities of pain and taste perception. Moreover, owing to the construct of the LMS, the median scores recorded in the LMS were lower than in the NRS (average pruritus: 3.3 [2.0–4.5] vs. 5.0 [4.0–7.0], P < 0.001; worst pruritus: 3.8 [2.6–5.3] vs. 7.0 [5.0–8.0], P < 0.001), which contributes to the reduction of the ceiling effect in the LMS. As such, the LMS is especially interesting for the use in clinical trials of severe CP indications.

The LMS assessing average pruritus showed a slightly better test-retest reliability compared with the LMS assessing worst pruritus. This may be explained by the longer intervals of the verbal anchors at the higher end of the scale. The average pruritus intensity is in most instances lower than the worst pruritus intensity and the higher concentration of verbal anchors at the lower end of the scale provides more orientation to patients, leading to better reliability scores. The excellent reliability of the LMS constitutes an advantage when comparing to the VAS, which performs less well in this regard. The lack of visual anchors in the VAS, which difficult the choosing of a precise score, may explain the worse reliability of this scale compared with the LMS. In our study, the reliability of the NRS was superior to that of the LMS. A better training of CP patients in the NRS may explain this finding. CP patients often associate their pruritus intensity with a number from 0 to 10, since this scale is regularly used in clinical practice either in routine questionnaires or verbally during the consultations. Furthermore, the numerical anchors of the NRS may suffice in guiding the patient to choose a precise score.

As for the convergent validity, we recorded stronger correlations between the LMS and the NRS (average pruritus: r = 0.808, worst pruritus: r = 0.783) compared with the correlations between the LMS and VAS (average pruritus: r = 0.570, worst pruritus: r = 0.535). Previous studies have shown strong correlations between the NRS and VAS. In our study, different recall periods used for the NRS (24 h as in the LMS) and the VAS (4 wk) may explain the observed discrepancies.

Patients with CP considered the LMS an appropriate tool and easy to use, with only a minority of patients manifesting difficulties in deciding on a score. A feasibility analysis from the perspective of attending physicians would also be desirable and should be considered in future studies.

The use of verbal anchors in the LMS constitutes one possible indication. The use of visual anchors in the LMS may also be desirable and should be considered in future studies.

In conclusion, the LMS is a suitable tool to assess CP with high acceptance from patients. Physicians should consider resorting to the LMS to assess possible variations in pruritus intensity in patients with severe CP.
Sources of funding

This study was supported by Beiersdorf AG.

Conflicts of interest statement

The authors declare that they have no financial conflict of interest with regard to the content of this report.

References

[1] Pereira MP, Stander S. Assessment of severity and burden of pruritus. Allergol Int 2017;66:3–7.
[2] Stander S, Zeidler C, Riepe C, et al. European EADV network on assessment of severity and burden of Pruritus (PruNet): first meeting on outcome tools. J Eur Acad Dermatol Venereol 2016;30:1144–7.
[3] Stander S, Augustin M, Reich A, et al. Pruritus assessment in clinical trials: consensus recommendations from the International Forum for the Study of Itch (IFSI) Special Interest Group Scoring Itch in Clinical Trials. Acta Derm Venereol 2013;93:509–14.
[4] Phan NQ, Blome C, Fritz F, et al. Assessment of pruritus intensity: prospective study on validity and reliability of the visual analogue scale, numerical rating scale and verbal rating scale in 471 patients with chronic pruritus. Acta Derm Venereol 2012;92:502–7.
[5] Reich A, Heisig M, Phan NQ, et al. Visual analogue scale: evaluation of the instrument for the assessment of pruritus. Acta Derm Venereol 2012;92:497–501.
[6] Hartmannsgruber MW, Swamidoss CP, Budde A, et al. A method for overcoming the ceiling effect of bounded pain scales. J Clin Monit Comput 1998;15:455–9.
[7] Parducci A, Wedell DH. The category effect with rating scales: number of categories, number of stimuli, and method of presentation. J Exp Psychol Hum Percept Perform 1986;12:496–7.
[8] Jenkins HH, Spencer ED, Weissgerber AJ, et al. Correlating an 11-point verbal numeric rating scale to a 4-point verbal rating scale in the measurement of pruritus. J Perianesth Nurs 2009;24:152–5.
[9] Green BG, Dalton P, Cowart B, et al. Evaluating the “Labeled Magnitude Scale” for measuring sensations of taste and smell. Chem Senses 1996;21:323–4.
[10] González-Fernández M, Ghosh N, Ellison T, et al. Moving beyond the limitations of the visual analog scale for measuring pain: novel use of the general labeled magnitude scale in a clinical setting. Am J Phys Med Rehabil 2014;93:75–81.
[11] Jones O, Schindler IC, Holle H. Assessing acute itch intensity: general labelled magnitude scale is more reliable than classic visual analogue scale. Acta Derm Venereol 2017;97:375–6.
[12] Bartoshuk LM. Comparing sensory experiences across individuals: recent psychophysical advances illuminate genetic variation in taste perception. Chem Senses 2000;25:447–60.
[13] Carlsson AM. Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. Pain 1983;16:87–101.
[14] Jang YH, Kim SM, Eun DH, et al. Validity and reliability of itch assessment scales for chronic pruritus in adults: a prospective multicenter study. J Am Acad Dermatol 2020;82:80–6.
[15] Rahbek O, Jensen SL, Lind M, et al. Inferior reliability of VAS scoring compared with International Society of the Knee reporting system for abstract assessment. Dan Med J 2017;64:A5346.
[16] Langner MD, Maibach HI. Pruritus measurement and treatment. Clin Exp Dermatol 2009;34:285–8.
[17] Hayes JE, Allen AL, Bennett SM. Direct comparison of the generalized Visual Analog Scale (gVAS) and general Labeled Magnitude Scale (gLMS). Food Qual Prefer 2013;28:36–44.