Assessing Acute Itch Intensity: General Labelled Magnitude Scale is More Reliable than Classic Visual Analogue Scale

Olivia JONES, Igor SCHINDLER and Henning HOLLE*
Psychology, School of Life Sciences, University of Hull, Cottingham Road, Hull HU6 7RX, U.K. *E-mail: h.holle@hull.ac.uk
Accepted Nov 17, 2016; Epub ahead of print Nov 21, 2016

The reliable measurement of itch intensity is crucial, both in research as well as clinical contexts. For example, when the reliability of a measurement scale is unknown, it is impossible to determine whether a patient has changed sufficiently to be confident that the change is beyond that which could be attributed to measurement error (1, 2). One factor that might influence the reliability of measurements is the type of rating scale used to assess itch intensity. Previous research (3–4) has documented the retest reliability of different rating scales for assessing chronic itch intensity. However, a retest reliability analysis of rating scales for acute experimental itch, induced using substances such as histamine or cowhage, is currently lacking.

Here, we compare the test–retest reliability of 3 rating scales commonly used for this purpose.

MATERIAL AND METHODS

First, we considered the visual analogue scale in its classic form (cVAS), where participants indicate itch intensity on a line ranging from 0 (no itch) to 100 (the most intense itch imaginable). Second, we included a variant of theVAS, where an additional ‘Scratch Threshold’ marker is set at 33% (tVAS [6]), defined as itching strong enough to be scratched (7). Finally, we considered the general Labelled Magnitude Scale (gLMS [8]), where participants judge the magnitude of itch on a line with quasi-logarithmically placed labels of “no sensation” at 0, “barely detectable” at 1, “weak” at 6, “moderate” at 17, “strong” at 35, “very strong” at 53 and “strongest imaginable sensation” at 100. Thus, all 3 scales have an identical range, but differ in the type and number of verbal labels provided (Fig. S11).

Ninety healthy volunteers took part after giving written informed consent. Twelve participants (gLMS group: n = 7, cVAS group: n = 5) were screened out as non-responders after the familiarization session (i.e., itch intensity ratings did not exceed 15) and one as an outlier (itch response above 3 SD of group mean), resulting in a final sample of 77 participants (38 females, mean age 24.66 ± 6.5; n = 25 in gLMS group, n = 26 in cVAS and tVAS group). Participants were told the study was investigating the experience of cowhage-induced itch and the scale. For this retest reliability analysis, we used a two-way mixed model, focusing on absolute agreement between sessions (10).

As shown in Table II, the gLMS had the highest retest reliability. This was the case regardless of which index was used to quantify itch intensity (peak: ICC 0.86; mean: ICC 0.71). The cVAS was the least reliable scale (peak: ICC 0.50; mean: ICC 0.45) and the tVAS had an intermediate reliability (peak: ICC 0.73; mean: ICC 0.64). Associated p-values, obtained using Fisher’s r-to-z transformation, indicated that the gLMS was significantly more reliable than the cVAS (p = 0.01, see Table II).

DISCUSSION

The higher retest reliability of the gLMS cannot be explained in terms of response clustering (i.e., the clustering of ratings around the verbal labels, see Appendix S11). Instead, our data suggest that retest reliability may be linked to the degree to which scales are open to interpre-

### Table II. Descriptive statistics of the two itch indices (mean, peak) for each session and scale group.

<table>
<thead>
<tr>
<th>Scale group</th>
<th>Index</th>
<th>Session 2 Mean (SD)</th>
<th>Session 3 Mean (SD)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>cVAS (n = 26)</td>
<td>Mean</td>
<td>24.38 (13.34)</td>
<td>27.45 (13.60)</td>
<td>1.11</td>
<td>0.28</td>
</tr>
<tr>
<td>Peak</td>
<td>64.92 (20.99)</td>
<td>66.31 (22.08)</td>
<td>0.32</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>tVAS (n = 26)</td>
<td>Mean</td>
<td>25.19 (12.05)</td>
<td>27.64 (13.78)</td>
<td>1.41</td>
<td>0.26</td>
</tr>
<tr>
<td>Peak</td>
<td>64.04 (21.20)</td>
<td>67.31 (24.54)</td>
<td>0.99</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>gLMS (n = 25)</td>
<td>Mean</td>
<td>16.24 (7.38)</td>
<td>18.59 (9.74)</td>
<td>1.83</td>
<td>0.08</td>
</tr>
<tr>
<td>Peak</td>
<td>48.12 (21.64)</td>
<td>49.68 (23.68)</td>
<td>0.65</td>
<td>0.52</td>
<td></td>
</tr>
</tbody>
</table>

c: classic; t: threshold; g: general.

---

1https://www.medicaljournals.se/acta/content/abstract/10.2340/00015555-2584

This is an open access article under the CC BY-NC license. www.medicaljournals.se/acta


doi: 10.2340/00015555-2584

Acta Derm Venereol 2017; 97: 375–376
tation. Previous research has highlighted that the lack of verbal anchors in the cVAS creates ambiguity, because participants are unsure where exactly they should place their mark (11, 12). This unsystematic variation may limit the reliability of the cVAS. In contrast, the tVAS adds a scratch threshold marker, providing participants with an additional landmark to guide their ratings which increases scale reliability. Finally, the gLMS with its 7 verbal anchors is least ambiguous and was found to be the most reliable scale for measuring acute itch.

Another factor that could explain the observed superior reliability of the gLMS is that this scale has been explicitly designed to yield ratio data, whereas it is strongly debated whether the cVAS provides ratio (13) or merely ordinal level data (for review, see 12). There is evidence that rather than providing a linear transformation of the internal representation of stimulus intensity, the cVAS provides only a non-linear representation, with a compression of scores especially at the top end of the scale (11). In contrast, the roughly logarithmic distance between the verbal anchors in the gLMS, determined in a semantic scaling procedure, has been demonstrated to yield ratio level data for ratings of oral sensations (14, 15) though a validation in the domain of itch is still outstanding.

A limitation of the present study is that participants were excluded from taking part in sessions 2 and 3 when their intensity ratings did not exceed 15 in the initial familiarization session. No participant in the tVAS group was excluded based on this criterion, but several in the gLMS (n = 7) and cVAS (n = 5) group, which may have biased the results. In general, obtaining very low ratings seems less likely when using the tVAS. Note, however, that this potential bias cannot explain the main finding of our study (gLMS is significantly more reliable in assessing peak itch than cVAS), since a comparable number of participants were excluded from these two groups.

In summary, our results suggest that the gLMS rating scale enables a more reliable measurement of acute itch intensity in healthy volunteers. The gLMS scale may be particularly suited for longitudinal studies, though care must be taken to avoid memory effects (e.g., by allowing for sufficient time between ratings, or by using distractor items). Since scale reliability is not a fixed property, but is also population-dependent (16), further studies are necessary to investigate whether these advantages of the gLMS scale generalise to experimental itch induced in chronic itch patients or to the clinical assessment of chronic itch intensity.

ACKNOWLEDGEMENT

Parts of this study were supported by a grant from the British Skin Foundation, awarded to HH (project number: 7011).

REFERENCES